National Library of Canada
Bibliothèque nationale du Canada

Canadian Theses Service
Service des thèses canadiennes

Ottawa, Canada
K1A 0N4

NOTICE

The quality of this microform is heavily dependent upon the quality of the original thesis submitted for microfilming. Every effort has been made to ensure the highest quality of reproduction possible.

If pages are missing, contact the university which granted the degree.

Some pages may have indistinct print especially if the original pages were typed with a poor typewriter ribbon or if the university sent us an inferior photocopy.

Reproduction in full or in part of this microform is governed by the Canadian Copyright Act, R.S.C. 1970, c. C-30, and subsequent amendments.

AVIS

La qualité de cette microforme dépend grandement de la qualité de la thèse soumise au microfilmage. Nous avons tout fait pour assurer une qualité supérieure de reproduction.

S'il manque des pages, veuillez communiquer avec l'université qui a conféré le grade.

La qualité d'impression de certaines pages peut laisser à désirer, surtout si les pages originales ont été dactylographiées à l'aide d'un ruban usé ou si l'université nous a fait parvenir une photocopie de qualité inférieure.

La reproduction, même partielle, de cette microforme est soumise à la Loi canadienne sur le droit d'auteur, SRC 1970, c. C-30, et ses amendements subséquents.

NL-339 (r. 88/04) c

Canada

University of Alberta

Staff Attitudes Toward Program Evaluation

By

Wayne Morris

A thesis submitted to the Faculty of Graduate Studies and Research in partial fulfillment of the requirements for the degree of Master of Education.

in

Vocational Education

Department of Adult, Career & Technology Education

Edmonton, Alberta

Fall 1991

National Library of Canada
Bibliothèque nationale du Canada

Canadian Theses Service
Service des thèses canadiennes

Ottawa, Canada
K1A 0N4

The author has granted an irrevocable non-exclusive licence allowing the National Library of Canada to reproduce, loan, distribute or sell copies of his/her thesis by any means and in any form or format, making this thesis available to interested persons.

L'auteur a accordé une licence irrévocable et non exclusive permettant à la Bibliothèque nationale du Canada de reproduire, prêter, distribuer ou vendre des copies de sa thèse de quelque manière et sous quelque forme que ce soit pour mettre des exemplaires de cette thèse à la disposition des personnes intéressées.

The author retains ownership of the copyright in his/her thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without his/her permission.

L'auteur conserve la propriété du droit d'auteur qui protège sa thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

ISBN 0-315-70277-X

Canada

University of Alberta

Release Form

Name of Author: Wayne Morris

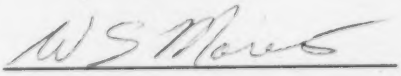
Title of Thesis: Staff Attitudes Toward Program Evaluation

Degree: Master of Education

Year This Degree Granted: 1991

Permission is hereby granted to the University of Alberta Library to reproduce single copies of this thesis and to lend or sell such copies for private, scholarly or scientific research purposes only.

The author reserves all publication and other rights in association with the copyright in the thesis, and except as hereinbefore provided neither the thesis nor any substantial portion thereof may be printed or otherwise reproduced in any material form whatever without the author's prior written permission.



Wayne Morris

79 Hawkwood Dr. N.W.

Calgary, Alberta.

T3G 3H9

October 4, 1991

University of Alberta

Faculty of Graduate Studies and Research

The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies and Research for acceptance, a thesis entitled Staff Attitudes Toward Program Evaluation submitted by Wayne Morris in partial fulfillment of the requirements for the degree of Master of Education.

Dr. D. Collett

Dr. D. Young

Dr. A. MacKay

September 26, 1991

Dedication

I wish to dedicate this doccument to my wonderful family, Katharine, Janelle and Heather. Had it not been for their patience, support and urging, this endeavor would never have been completed.

Abstract

This study undertook to determine the attitudes of staff within the Calgary Board of Education, towards program evaluation. Specifically, the extent of involvement and the benefits and drawbacks were assessed. A proportional random sampling procedure was used. A survey method was used to collect the data from the participants. The instrument was created using attitude statements derived primarily from Turner and Clift (1985). The statements were rated on a Likert-type five point scale with strongly agree at one end and strongly disagree at the other. The data were compiled and mean values were used to represent the group attitude. Subgroups of experience, gender and school type were used to determine differences in attitude. Analysis of variance was used to determine the statistical significance of the differences at a level of .05. Factor analysis was used to compare these attitudes with those found to exist in the Solihull study by Turner and Clift (1985).

Two basic issues were used to determine attitude differences; evaluation control and formality. This created three subclassifications of evaluation; informal, formal-internal, and formal external. The results show that staff involvement with program evaluation has been diverse. Involvement in informal evaluation is greater than the other subclassifications. The least experienced group of teachers had the highest percentage involvement with informal evaluation. Eight items seen as benefits were generally determined to be so across all three subclassifications. As formality increased and control moved to external, these benefits were not rated as strongly. Items shown to be disadvantages involved potentially threatening situations for teachers. The use of student scores from provincial tests and administrators only evaluating programs, were the major items of drawback in all subclassifications. Staff generally associate internal evaluation with program

improvement and external evaluation with accountability. Staff also want to know that improvement will take place as a result of being involved in program evaluation.

Staff attitudes generally fell into the same conceptual framework proposed by Turner and Clift (1985) for formal-internal evaluation. The concepts of collegiality and openness, professionality, formality, efficiency and threat and professional development were all found to be components of these staff attitudes for formal-internal evaluation. When informal evaluation attitudes were compared with those of the Solihull study, one general difference existed pointing to the possibility of a different major factor being involved, which may be "practicality".

Table of Contents

Chapter		Page
1.	Orientation to the Problem	1
	Problem Statement	3
	Subproblems and Objectives	4
	Definition of Terms	5
	Significance of the Study	6
	Scope of the Study	9
	Delimitations	9
	Limitations	9
	Assumptions	10
2.	Overview of Program Evaluation	11
	Evaluation: A Tool for Excellence	13
	Evaluation for Attaining Focus on the Total School	14
	Evaluation for Program Improvement	17
	Control of Evaluation	20
	External	20
	Internal	22
	Existing Procedures	24
	Problems and Promises	31
	Summary	36
	Conclusion	38
3.	Design and Methodology	40
	Population and Sampling	40
	Confidentiality	40
	Data Collection Techniques	41
	Research Instrument	42

Chapter Page

The Data ... 43

Subproblem 1 ... 43

The data needed ... 44

How the data were screened ... 44

Analysis of data ... 44

Interpretation of the data ... 44

Subproblem 2 ... 45

The data needed ... 45

The means of obtaining the data ... 45

How the data were screened ... 45

Analysis of data ... 46

Interpretation of the data ... 46

Subproblem 3 ... 46

The data needed ... 47

How the data were screened ... 47

Analysis of data ... 47

Interpretation of the data ... 47

Subproblem 4 ... 48

The data needed ... 48

How the data were screened ... 48

Analysis of data ... 49

Interpretation of the data ... 50

4. Presentation and Analysis of Results ... 51

Sample Demographics ... 51

Extent of Involvement in Program Evaluation ... 53

Benefits and Drawbacks ... 58

Chapter Page

Informal Program Evaluation 60
Informal Program Evaluation Subgroup Differences 62
Teaching Experience 62
Gender 63
School Type 65
Formal-Internal Program Evaluation 65
Formal-Internal Program Evaluation Subgroup Differences 67
Teaching Experience 68
Gender 69
School Type 70
Formal-External Program Evaluation 70
Formal External Program Evaluation Subgroup Differences 73
Teacher Experience 73
Gender 73
School Type 75
Comparison of Means by Evaluation Subclassification 77
General Impressions of Program Evaluation 80
General Impressions Subgroup Differences 81
Teaching Experience 81
Gender 82
School Type 82
Attitude Categories 83
Formal-Internal Evaluation Factors 84

Chapter Page

Informal Program Evaluation Factors 88

5. Summaries, Conclusions, Implications and Recommendations ... 93

Summaries and Conclusions 93

Subproblem One - Summary 93

Informal program evaluation. 94

Formal-internal program evaluation. 95

Formal-external program evaluation. 95

Subproblem One - Conclusions 96

Subproblem Two - Summary 96

Informal program evaluation. 97

Formal-internal program evaluation. 98

Formal-external program evaluation 99

Subproblem Two - Conclusions 103

Subproblem Three - Summary and Conclusions .. 104

Subproblem Four - Summary 106

Subproblem Four - Conclusions 107

Implications and Recommendations 109

Educational Implications 109

Research Implications 112

Recommendations 113

Education 113

Research 114

Bibliography 116

Appendix 1. Instrument Cover Letter 120

Appendix 2. Survey Instrument 122

Appendix 3. Rotated Factor Loadings 129

List of Tables

		Page
Table 1	Distribution of Participants	51
Table 2	Distribution of Respondents by School Type	52
Table 3	Extent of Involvement in Program Evaluation by Classification	54
Table 4	Distribution of Involvement in General Program Evaluation by Years of Teaching Experience	55
Table 5	Distribution of Involvement in Informal Program Evaluation by Years of Teaching Experience	56
Table 6	Extent of Involvement in Formal-Internal Program Evaluation by Teaching Experience	57
Table 7	Extent of Involvement in Formal-Internal Program Evaluation by School Type	58
Table 8	Extent of Involvement in Formal-External Program Evaluation by Teaching Experience	59
Table 9	Extent of Involvement in Formal-External Program Evaluation by School Type	59
Table 10	Mean ratings for Informal Evaluation	61
Table 11	Means of Data: Informal Evaluation Grouped by Teaching Experience	63
Table 12	Means of Data: Informal Evaluation Grouped by Gender	64
Table 13	Means of Data: Informal Evaluation Grouped by School Type	66
Table 14	Mean Ratings for Formal-Internal Evaluation	67
Table 15	Means of Data: Formal - Internal Evaluation Grouped by Teaching Experience	68

Page

Table 16 Means of Data: Formal - Internal Evaluation Grouped by Gender 69

Table 17 Means of Data: Formal - Internal Evaluation Grouped by School Type 71

Table 18 Mean Ratings for Formal-External Evaluation 72

Table 19 Means of Data: Formal - External Evaluation Grouped by Teaching Experience 74

Table 20 Means of Data: Formal - External Evaluation Grouped by Gender 75

Table 21 Means of Data: Formal - External Evaluation Grouped by School Type 76

Table 22 Comparison of Means by Subclassification of Evaluation Type 78

Table 23 Means - General Impressions of Program Evaluation 81

Table 24 Means - General Impressions of Program Evaluation Grouped by Teaching Experience 82

Table 25 Means - General Impressions of Program Evaluation Grouped by Gender 83

Table 26 Means - General Impressions of Program Evaluation Grouped by School Type 84

Table 27 Factor One: Professionality / Professional Development 85

Table 28 Factor Two: Efficiency 86

Table 29 Factor Three: Threat 86

Table 30 Factor Four: Collegiality / Openness 86

Table 31 Factor Five: Formality 87

Page

Table 32 Factor One: Efficiency / Openness 89

Table 33 Factor Two: Threat 89

Table 34 Factor Three: Formality 90

Table 35 Factor Four: Professionality 90

Table 36 Factor Five: Professional Development / Growth 91

Table 37 Factor Six: Collegiality 92

Table 38 Percentage Involvement in Informal Program Evaluation by Teaching Experience 94

Table 39 Percentage Involvement in Formal-Internal Program Evaluation by Teaching Experience 95

Table 40 Benefits and Drawbacks Compared by Evaluation Subclassification 102

Chapter One

Orientation to the Problem

Program evaluation as a field of study has been around since the 60's, but had very little direct impact on schools and school programs. Extensive research has been conducted into various aspects of why it has failed to meet its goal of providing "relevant, useful, timely information on the value of educational and social programs" (Smith, 1981a, p. 79). Herman provides us with a simple summary of the major reasons that program evaluation is generally limited to being used for accountability.

> Evaluation has been primarily linked with 'top-down,' highly centralized improvement approaches which were not necessarily sensitive to 'bottom-up' needs; evaluation data has been derived primarily from tests of student achievement which examine only a narrow range of outcomes; evaluation often ignores critical variables in the context and process of schooling; evaluations have not sufficiently considered the factors which would facilitate attention to findings and translation of findings into action. (Herman, 1987, p.4)

Schools have been under pressure in recent years to increase their effectiveness, while at the same time reduce the costs of providing service. Improvement strategies first focused on effective teaching methods and enhanced curriculum, moved to effective leadership, and are now beginning to focus on program effectiveness.

As part of the focus on program effectiveness, the curriculum teams within the Calgary Board of Education have developed a series of Indicators of Effective Programs (1989). These indicators are to be used either by individual teachers or groups of staff members as a guide for ongoing program development and improvement. One of the "indicators" is program evaluation. "Program evaluation

is an ongoing, planned activity which ensures that curricular objectives and student needs are met" (p. 9). Orientation to the rationale for the development and utilization of these indicators is gradually filtering down the infra-structure, to the school administration. The methods which are currently used for program evaluation are as varied as the schools and programs involved. No specific methods, timelines or guidelines for evaluating programs have been established, although three points are suggested to be included when establishing a procedure. These include using a variety of evaluation strategies, including students, teachers, and both school and non-school based personnel, as well as providing for community input.

This approach intends staff to be introspective of the programs offered in their schools. It further suggests that staff can improve programs through some form of "internal evaluation". The indicators which are provided are then interpreted to be the basic standards by which effective programs should be judged.

The introduction of the internal evaluation approach is not new, and has met with mixed results in many countries. Love (1983) states that

> internal evaluation systems are intended to influence decision making and organization performance in two major ways: first, by providing relevant and timely information on which managers can base decisions; second, through the evaluation process itself, which has certain effects on decisions and performance as well as on the decision-making process. (p. 8)

It is not just the evaluation that is important, but also the professional growth which Smith (1983) states has always been considered a by-product of the evaluation process. The personal involvement can become the impetus for change. A focus on wanting to improve the program becomes the issue, rather than maintenance of the status quo.

Turner and Clift (1985) have determined five categories of attitudes to be involved in school self-evaluation: collegiality and openness, professionality, formality, efficiency and threat and professional development. It is important to understand if these same categories of attitudes exist within the Calgary Board of Education. From an academic view, it is important to determine the existence of such categories within this population, as a start to better understanding the relationship between these aspects of eduction. The Solihull study (Turner and Clift) was the only other research of similar nature that has been published up to the end of 1989. The reoccurrence of these categories would provide some indication that the phenomenon could be generalized to a larger population. The answer could help in the creation of an overall strategic approach to meeting the needs of students through planned and funded change.

It is evident that there is substantial interest in school-based evaluation, and much research which points to its characteristics as important in the utilization of evaluation information. If ongoing internal evaluation at the school level is to be expected or encouraged, for the purposes of either program improvement or accountability, then further research is needed to determine the staff attitudes surrounding program evaluation. This added information could provide a background upon which to design an implementation plan. Without this information, inaccurate presumptions may be made about the intent with which staff evaluate programs.

Problem Statement

What has been the extent of Calgary Board of Education staff involvement in the evaluation of school programs, and what are the perceived benefits and drawbacks for being involved in the process?

Subproblems and Objectives

1) To what extent have staff been involved in "evaluating" school programs, either formally or informally?

a) Determine what proportion of staff have been involved in the evaluation of programs.

b) Categorize the type of evaluation staff have participated in as either informal or formal with internal or external control.

c) Determine if similarities or differences by subgroup exist.

2) What are the perceived benefits and drawbacks of program evaluation?

a) Evaluate the views of all respondents, as to the perceived benefits and drawbacks of the evaluation types which they have experienced.

b) Determine the differences which exist between different categories of program evaluation.

c) Assess the mean attitudes of staff within the following groups which make up the population: school type, gender groups and experience groups.

d) Determine if there is a difference in attitude among the subgroups.

3) Determine the views of all respondents on selected issues relating to the benefits and drawbacks of implementing program evaluation.

a) Assess the mean attitudes of staff as a whole for each issue.

b) Determine the mean attitudes for each issue, grouped by school type, gender and experience.

c) Determine if there is a difference in attitude among the subgroups for each issue.

4) To what extent do the attitudes of staff fall into the categories determined to exist by Turner and Clift (1985) in a similar study of schools in Solihull England?

a) Duplicate the process of factor analysis undertaken by Turner and Clift (1985).

b) Make corrections for terminology between England and Canada.

c) Compare and contrast the results.

Definition of Terms

The following terms used in this study may be understood in a variety of ways. For the purpose of clarity and interpretation, their use is defined.

Program: The term "program", when used in the school context, usually refers to "a planned sequence of activities intended to achieve some goal" (Meyers, 1981, p.1).

Evaluation: Stufflebeam defines evaluation as "The process of delineating, obtaining and providing useful information for judging decision alternatives" (1971, p.40).

Formal participation in evaluation is hinged on the definition of formal. The Heritage Dictionary defines **formal** as "Following or adhering to accepted forms, conventions or regulations" (1975, p. 516). Therefore, formal participation in evaluation is defined as adhering to accepted procedures, processes or conventions.

Informal participation in evaluation is defined as teacher initiated, involving only the teacher and / or the classes involved, and following no formalized, structured or accepted procedure.

School staff: **School staff** as used in this study refers to administrators and teachers, who perform either classroom or support duties, all of whom must be school-based.

Internal: Within the context of the organization, in this case the school, using the resources available within its' walls.

Significance of the Study

Much research has been done in all aspects of program evaluation, as it pertains to all publicly funded programs, not just education. This activity has been driven by the demand for public accountability for funds spent, and the desire to receive quality results. Within the field of education, there has been repeated call for higher standards, and the pursuit of excellence. Indeed with finite and shrinking

budgets, the quest for increased effectiveness of schools, teachers and classrooms has become essential.

Many of the attempts to improve the product were aimed at either fine tuning or drastically altering the individual components of education. Change has now become the one constant. Society was changing not only in its fundamental unit of the family, but also in its way of doing business. Public education as a system now had to adapt to changes in the fundamental belief and values system in which the child was being raised. At the output end of the system, drastic changes were envoked to create workers who had the characteristics needed for an evolving workplace. A greater proportion of the transitionary economy was becoming based upon the servicing of information needs, rather than the production of goods. This called for skills of thinking and analysis at the higher order of Bloomes taxonomy. Society was rapidly becoming a global village, where the need for international awareness was as important as community awareness.

The use of internal evaluation as a tool in education, for meeting the constantly changing needs is a move towards greater local autonomy and control over education. But as Love (1983) has stated, internal evaluation has two components inherent in the process. One is the information gathered, and second is the focus on improvement which the staff gain from participation in the process. The quality and extent of these will be affected by the attitudes which the staff have to participation in the process. This provides a significant reason for undertaking a study of staff attitudes toward program evaluation.

This study is also important because the term "evaluation" has historically been associated with a degree of fear. Love (1983) describes organizational response to evaluation as "characterized by anxiety, avoidance, immobilization, fear of failure

and similar dysfunctional behaviors" (p. 8). The introduction of internal program evaluation for the purposes of improvement of service provided must be based on the assumption that the staff are professional and would carry out an evaluation honestly, thoughtfully and carefully. If that staff respond with the characteristics provided by Love, the process will be a negative step in the development of education. The attitudes of staff will partially determine the success or failure of this scheme for program improvement.

Research tells us that the utilization of evaluation information is related to the degree to which staff is involved in its generation. Involvement in the process of evaluation will take time away from other activities which teachers are normally involved in. Will the time be excessive? Will teachers resent the effort? Will they find the recommendations useful or timely? It is questions such as these which this study attempts to answer. A description of staff attitudes towards program evaluation has not been done within the Calgary Board of Education. The results of such a study should be of great interest to senior decision makers.

While wanting to keep up with the current trends, schools are obligated to utilize their resources wisely, maximizing the educational benefit for individual student learning. Actual practice in schools reveals that decisions are often made based on limited information, or more often, a conscious decision is made not to change the existing program. This practice leads to stagnant education. Schools require methods for evaluating their practice, with the intent of making better decisions about changes that will improve the educational service they provide for their clients. The program improvement and long range planning which will enable enlightened decisions to be based on the best, and most complete information which is available.

Scope of the Study

Delimitations

Program evaluation may be considered by some as inseparable from teacher evaluation. For the purposes of this study, program evaluation is not to involve the evaluation of teachers.

This study is also delimited to the determination of attitudes and perceptions of school staff specifically related to a) staff experience with program evaluation either formally or informally and; b) determining the perceived benefits or drawbacks of evaluating programs. (This research will utilize the five categories which Turner and Clift (1985) derived using factor analysis in a series of studies in Britain.)

Limitations

The instrument used to collect data was carefully constructed using questions which Turner and Clift (1985) used in their Solihull study. Some statements were added, and some British terms were modified with language more appropriate to Calgary. To collect data which reflects participant involvement in the three basic forms of program evaluation, and responses of a general nature notwithstanding evaluation experience, the instrument was prepared in four sections. This represents another departure from the format used in the Solihull study. This study is then limited by the ability of the instrument to adequately measure the true attitudes and opinions of the participants. Being that the instrument has not been standardized, it may subject the findings to some question.

Other factors which may tend to weaken the generalization of the results to the population involve limited sample size, collection of the data within the month of June and response rate.

Assumptions

The study assumes that the respondents in the selected sample were able to provide a valid estimation of their opinions regarding participating in the evaluation of school programs, given definitions, examples, and a brief explanation of the intent..

A second assumption underlying this study was that respondents would give serious and careful consideration to what is being asked.

A third assumption was that the questionnaire presenting selected benefits/drawbacks and methodological statements is a valid method for gathering perceptions about participating in the evaluation of school programs.

Chapter Two

Overview of Program Evaluation

The evaluation of programs is not a new phenomenon, but the formal study of methods and practice as a formal discipline is. This discipline known either as program evaluation, or evaluation research has undergone a number of changes during the short time since its formal inception in the early '60s. There has been great debate amongst its scholars, over the purpose, meaning and methods of program evaluation. According to Brewer (1983), "some of these controversies reflect fundamental differences among evaluation researchers and utilizers . . . differences that reflect the diverse origins of the field" (p.16). Those "origins" of program evaluation, which Brewer refers to, come from the disciplines of psychology, education and economics. The various definitions generated from this diversified origin of program evaluation, all have as their central theme the assessment of value, worth or merit of a program. Stufflebeam (1971) broadened the view by defining evaluation as "the process of delineating, obtaining and providing useful information for judging decision alternatives" (p. 40). Although the debate continues to rage over the definition of terms, it is Stufflebeam's definition which appears most often in evaluation literature, and therefore seems to be generally well accepted.

Having achieved some general consensus about the definition, controversy continues over the process or methods which are most appropriate to delineate and obtain information about the program. With each model that is conceptualized, the question of outcome or product validity is raised. The arguments are numerous and far too diverse to summarize here, but some common themes have emerged. From one side, it has been generally well accepted that the scientific model of evaluation is well suited to the evaluation of anything, educational programs being no exception.

(Rossi & Wright, 1977, p.13) The argument is based on the fact that the scientific model utilizes the randomized controlled experiment, which promises greater internal and external validity. In recent years this particular view has been criticized by educators and researchers alike, because in their view the experiment cannot always be properly implemented. This prompted researchers to promote a more holistic, naturalistic, or humanistic approach to evaluation. One of the fundamental arguments put forward is that the scientific method is insensitive to local conditions, due to its dependency upon large sample statistics. (Parlett & Hamilton, 1976; Guba, 1978) A second argument questions whether evaluation is "primarily an inquiry-based social service designed to serve clients, or a field-based research activity, designed to advance knowledge" (Smith, 1981, p. 21). This question posed by Smith, outlines the dichotomy between evaluation as a research tool for the generation of knowledge and evaluation as a tool for pointing out the strengths and weaknesses of programs. Methods which are research oriented tend to use the scientific model while others which are intent on assisting program staff or administration improve programs, choose alternate methods. Herman (1987) suggests that the "clients" of evaluations are not generally school-based, and as a consequence, schools typically cannot utilize the results.

> Evaluation has been primarily linked with 'top-down,' highly centralized improvement approaches which were not necessarily sensitive to 'bottom-up' needs; evaluation data has been derived primarily from tests of student achievement which examine only a narrow range of outcomes; evaluation often ignores critical variables in the context and process of schooling; evaluations have not sufficiently considered the factors which would facilitate attention to findings and translation of findings into action. (Herman, 1987, p.4)

In the last ten years, serious consideration has been given to the utilization of evaluative information by the client, providing another argument for moving away

from the traditional scientific methods, towards techniques which provide greater insight and therefore are more useful to the clients. Part of the reason for these arguments stems from the belief that traditional models of program evaluation have not lived up to their promise of providing "relevant, useful, timely information on the value of educational and social programs" (Smith, 1981a, p. 79). Smith is not alone in his conviction that traditional program evaluation models need revision for use in education today. Many authors have written expressing those same views, calling for methodologies which will better meet the needs of education. (Stake, 1975; Ayers, 1986; Herman, 1987) The major obstacle to the development of a single model is that education has so many different needs, no one model can possibly serve all its purposes. If evaluation in education is ever to live up to its promises, a series of models and methods are required, each being tailored to suit a specific need. No one perfect model, which meets all the needs for the evaluation in education exists, as the history of program evaluation has shown. The proliferation of new and unique models over the last 10 years has been an attempt by researchers, to look differently at evaluation, through individualizing models to suit some of the specific needs.

Evaluation: A Tool for Excellence

The report A Nation at Risk: The Imperative for Educational Reform (National Commission on Excellence in Education, 1983) which exposed the state of education in the United States, and subsequent similar reports in Canadian education have brought about an era where educational systems are in search of effectiveness or "excellence". A great deal of research has been accumulated on the topic of "what makes schools effective". The core of this research contains indicators of both effective schools and effective programs. A multitude of different strategies for achieving "excellence" have been developed around these indicators,

and are currently being implemented in almost every school nation wide. Alberta Education has responded to the call for excellence by revising the curriculum and changing the graduation requirements. These changes were designed to better prepare students to meet the demands of society. The pressure for schools to improve their programs, and move towards "excellence" is very real.

But how do schools know if they are approaching excellence in educational service? "Excellence" is a utopian construct which cannot be simply operationally defined, but implies better than either average or good. Because "excellence in education", is a construct which must be generated within a philosophic frame of reference, attainment is possible only after specific school-based definition. Measures of educational excellence are too often defined external to the school, such as the use of achievement scores (Herman, 1987), but this vision of educational excellence fails to describe the richness of an excellent educational experience. Because of this difficulty, emphasis has been placed on improvement; a shift from competence towards "excellence". This shift implies the necessity for ongoing assessment of educational practice. Often, this assessment centers around the utilization of indicators which have been prepared by program experts. One of the indicators of an effective program, as set out by the Calgary Board of Education, is evidence of ongoing program evaluation. (Calgary Board of Education, 1988, p. 9) Any model or method which is designed for the evaluation of educational programs must emphasize utility of the evaluative information in the improvement process of the school, thereby assisting in the movement toward "excellence".

Evaluation for Attaining Focus on the Total School

One major premise upon which this paper is based, deals with the school as a microcosm of society, with its own culture, beliefs, values and standards. The school

is also viewed as the "molecule" or the fundamental building block of general education. The success or failure of the school in its mission determines the outcome of education. The "structure" of the total school program is comprised of many "atoms"; teacher - student and student - student interactions, along with curricular and extra curricular programs which combine to create the unique character, educational climate and total program of education within the building. This view of education is supported by most educators and illustrated by DeRoche's statement that "the basic, most fundamental element of education is the interaction between a group of students and their teacher" (1987, p.15).

The term "program", when used in the school context, usually refers to "a planned sequence of activities intended to achieve some goal" (Meyers, 1981, p.1). Therefore, the goals of education to which a school is committed, determine the "programs" which are implemented. Because education has historically been compartmentalized into the study of subjects which society values, and the study of these subjects has in most people's minds become the end or goal, then it is easy to understand why most subjects are referred to as programs. Programs, from the school perspective, are most often the implementation of subject based curricula as prescribed by the government. This implementation varies widely due to factors such as teacher expertise, resources, and student needs.

But curricular programs are not the only programs which schools offer students. Schools have traditionally provided students with clubs and other activities to supplement and enhance the academics. Today with schools striving for "excellence", many special programs are being implemented to deal with a multitude of topics such as individual student needs, libraries, leadership, computers, and others which cut across traditional curriculum lines. These programs, which are designed to provide important benefits to students, can easily

get lost in the compartmentalized school where teachers focus all their attention on specific curricula or an individual grade. Integrated programs require a strong commitment from the school staff to achieve success and maintain it. Part of the reason is that integrated programs are usually an add on to the current work load of the teacher. For the staff to contribute the effort which is required to bring integrated programs to their fullest potential, they must understand and appreciate what their extra efforts are doing for students. This understanding and appreciation is more meaningful in the context of the total school program. For schools to achieve true success or excellence, the staff must focus more often on the total school, or the collective contribution which their programs make toward student education. This requires processes which assist staff with the construction of the total school view, or team approach to education. One such process should involve a method of program evaluation, through which the educational team gains a better understanding of the contribution which each program makes to the total education which students receive.

The key players on the team of educational process, who together build and nurture the educational climate are the teachers, school based administration, parents and students. For the school to attain success in its mission, and move towards excellence, the key players must all work together as a team. If the school operates as a team, it then follows that any evaluation of programs must encompass all the members of that team in order to gain a true perspective of the wide ranging contribution individual programs have on the total school program. The inclusion of all parties that have a vested interest in the educational process is called stakeholder participation. Meaningful stakeholder participation as defined by Greene (1988) has the stakeholders involved in "shared decision making" (p.101). To become truly involved in the decision making process, Greene describes the

stakeholders "not as advisers or consultants, but as active [and] engaged" (p.101) or to use Stake's words "collaborators in inquiry" (1983, p.18). Alkin, Daillak and White (1979) as well as Greene (1988) agree that stakeholder participation in evaluation is an important key to utilizing the results of the evaluation, and increasing participants' readiness for the change process.

Evaluation for Program Improvement

The work of Windele and Neigher (1978) provides us with the insight that evaluative information can be used for three main purposes, accountability, advocacy and program improvement. It would be extremely difficult for any one evaluation to adequately meet all three of these goals at the same time. This is where the traditional vision of evaluation fails in its lofty aspirations. All three uses require information which differs in some way. A competent evaluation must have a focus or a goal, in order to be successful in meeting its obligations for appropriate information. Typically, evaluations for accountability and advocacy are not great at helping schools improve their programs. Conversely, evaluations for program improvement are not terrific for public accountability or advocacy.

Schools are guided by a set of criteria dictated by either their board of trustees, their government or both, which delineate the basic goals of education, and the curriculum through which these goals should be reached. Accountability of schools to their boards and governments is often a question of whether these goals are being reached, and to what extent. The methods, both formal and informal, often include the use of quantitative measurements of student achievement based on the prescribed curriculum. (DeRoche, 1987) External evaluation monitors, as impartial

bodies, are well suited to provide a greater measure of credibility for accountability with government or the community but traditional comprehensive evaluations are rarely used on single schools.

Advocacy is usually a process by which schools boast of how well they meet the objectives, when compared to other schools in the jurisdiction. The purpose is usually self-serving, with primary motivation being either prestige or the attraction of more clients. Neither of these two purposes are considered to be as important for schools, as is the third, program improvement.

The crucial factor remains that the school's needs for information to assist in program improvement, cannot be adequately met through the evaluative methods used for accountability or advocacy. Rein (1981) suggests that program improvement or progress toward excellence, may be approached by the school or individual, through the study of the program. "Experience in implementing a program often leads to a change of mind about what the program can or should do. A review of outcomes helps in understanding intentions" (Rein, 1981, p.132). Rein further states that

> a useful evaluation essentially requires learning about the experience of the program from two perspectives: why the program did what it did and what outcomes it achieved by taking account of a broad range of measures that reflect different interests. (p. 133)

Program improvement embodies the reassessment of the philosophy and goals upon which the program was based or implemented, as well as the "what and how" or the methods used in the school to meet the goals of education through curricular and extra curricular means. To evaluate the program requires a method which is sensitive not only to the individual school setting, but also the clients of that school.

Meyers (1981) defines program evaluation as " the effort to understand the functioning and effects of a program, which is a planned sequence of activities intended to achieve some goal" (p.1). This understanding of the program which Meyers refers to requires a very close contextual link with the school. This is because understanding is created through a process of first hand experience and meaningful discussion. If this meaning is created without thoroughly being immersed in the cultural environment which implemented the program, the meaning may not retain the context, and lose both credibility and utility for the school. This contextualization is necessary because schools vary widely in the beliefs they hold, and the approaches which they use in reaching those global goals of education. Program evaluation for improvement purposes must focus on the "how and what" of educational process, which is closely connected to the (why) beliefs and values held in within a school. Proper understanding cannot be gained outside the school context.

Because schools develop characteristics and climates which are the embodiment of their leadership, many different school environments exist. There are schools which have an interest in reflecting on their current practice, for the purpose of refining and improving education. As well, there are schools which strive to maintain the same program, because it has always served them well. The purpose of program evaluation would be best served, when there is a strong need for the information which the evaluation can provide. Mandatory improvement oriented evaluations would not be worth the effort.

Control of Evaluation

External

The major models available for program evaluation, primarily use an external team of evaluators to study the institution. The reasons for external control of evaluations vary from objectivity, to methodological expertise, to sheer work load. From the school perspective, the use of an external evaluator has many drawbacks. The first deals with context. For evaluation to maintain the context of the school, the evaluation would be better suited to an internal evaluator. Campbell explains that

> External evaluators tend to lack the essential qualitative knowledge of what happened. The chronic conflict between evaluators and implementors, which will be bad enough under unified local direction, tends to be exacerbated . . . the relevance of the measures to local program goals and dangers is weakened (1979, p. 71).

For an external evaluator to "soak up" the context of the school, means leaving behind his preordinate evaluation style, where the intent is to measure the extent to which objectives were attained. This usually requires more time than the evaluator would be willing to put in.

Among the major models of program evaluation which are currently applicable to school based evaluation, is Stake's model of responsive evaluation. This model is based on the client centered approach, and tries to respond to the needs for information which the various stakeholder groups may have. Very simply, Stake's approach compares the intents of the program with what is observed. In doing so it emphasizes the "setting" in which the learning takes place, coming close to alleviating the problem of context. This is primarily accomplished through the continual involvement of stakeholders. Stake's model of responsive evaluation

is the product of strong beliefs about what education is and should be. Stufflebeam and Shinkfield's description of Stakes' philosophy of education very much parallels the assumptions about education made in this paper.

> He believes that education is organic, not produced. He sees all people striving to excel - but not necessarily at the education the state prescribes. He believes that formal education changes informal education very little, that context counts most, and that teachers are more a part of context than creators of education. He said that goals of education cannot be other than the variation of aims embodied in the ideals and convictions of different communities, families, and students. He strongly favors local autonomy in the basic governance of schools. (1985, pp. 239-240)

Understanding Stake's educational philosophy, adds greater appreciation for the intention of his evaluation model. The use of this model for program improvement purposes is very appealing for schools, in that strengths and weakness of the program are pinpointed. Responsive evaluation is oriented towards having the clients (stakeholders) determine those things which require investigation or the questions which should guide the investigation. To that end, it allows the teachers, administration and parents to choose what they require more information about, and the direction they wish the evaluation to take. The major drawback of Stake's model is the second major drawback of external evaluations. It is the inability of schools to afford ongoing external evaluation. Externally controlled evaluation for the purposes of improving school programs is not feasible for local or individual school use. The cost to education in dollars would be prohibitive. With the focus for evaluation, shifting to the school and the types of programs which it offers, many of the traditional models of program evaluation lose their usefulness. Several new models of evaluation for use by schools have been devised around self-evaluation.

Internal

Evaluations which are devised around institutional self-introspection are not new. "The notion of the self-evaluating organization that uses program evaluation as the basis for program development and change . . . was described by Aaron Wildavsky" in 1972 (Love, 1983, p.5). Institutional self-evaluation is described by Smith (1981) as evaluation which is carried out by one piece of a large organizational enterprise. (p.82) Clift, Nuttall and McCormick, (1987) refer to their experiences with institutional self-evaluation as

> whole school evaluation although . . . not necessarily . . . all staff and all areas of the curriculum were evaluated, but that the activity was visible as a school activity. Activities that were initially department- or classroom-based, or concerned an issue that was 'institution-wide', were included if they formed part of the whole school programme. (p. 175)

Self or internal evaluations have been used for a few years in both higher education and industry, in Britain, Sweden, the United States and Canada. Each country has created their own terminology to describe the process, and refer to it as any one of: self-study, internal evaluation, activity evaluation or institutional self-evaluation. In this paper, all of the terms above are considered to be synonymous. The common factors in all these methods involve the lack of an external evaluator, and the utilization of the results for the improvement of programs. The most common reason for the existence of each approach appears to be the belief that there are benefits to be gained by the institution through the internal evaluation process.

Attkisson and Broskowski (1978) believe that program evaluation is a tool to serve the needs of the organization, for decision making and planning. Love (1983) takes that one step further by noting that

> internal evaluation systems are intended to influence decision making and organizational performance in two major ways: first, by providing relevant and timely information on which managers can base decisions; second, through the evaluation process itself, which has certain effects on decisions and performance as well as the decision-making process. (p.8)

If the information which the evaluation provides is valid and meaningful to the institution, then it follows that evaluations can have an effect on decision making. At the school level, decisions are often a cooperative venture, with input and participation by teachers and administrators. Internal program evaluation could play an important part in movement of schools from competence towards excellence by providing useful information for improvement. But it is the effects of internal evaluation process which can also contribute substantially to the improvement of the program or school, through the reassessment of beliefs and goals, a focus on the total school program, and a clearer understanding of the program.

Ewell (1984), in his work on self-evaluation programs for higher education, makes a statement which is equally applicable to schools as well.

> Educational institutions, like other collective enterprises, cannot hope to achieve excellence unless they can effectively communicate to others the particular business they are in. Doing so requires development of a common language centered not around the resources they possess or the reputation they have achieved, but on what they, in fact, expect and hope to achieve. (p. 16)

This statement drives home the importance of schools grounding their practice in the educational philosophy and beliefs which are commonly held by the staff. This must be done in such a way that what they believe becomes part of the common language which binds the staff into a team. The programs the team offer must then be guided by goals or objectives, which are a natural outcome of what the team believes their practice is all about. Evaluation of either the total school program or individual programs offered by the school must involve a reassessment of not only the goals of the program, but the beliefs upon which the goals were based.

A second point derived from Ewell's (1984) statement, is that the total staff must understand the program which is offered. This can be a function of the evaluation process itself, with the first act of evaluation being the description of the program. In terms of responsive evaluation, this description would be of the program intents. Then as the data collection and analysis continues, and the observations of the program are compared to the intents, not only will the strengths and weakness be elaborated, but a much greater understanding of the program itself will be generated.

Existing Procedures

Several states such as California and Ohio have adopted the accreditation model of evaluation, to fulfill both the needs for accountability and improvement. The process requires the school to perform a self-study (internal self-evaluation), which is directed by either a single person or a committee comprised of staff members. Once this evaluation has been completed, it is followed up by a visitation from an external accrediting team. The purpose of the external team is to validate or refute the self-study performed by the school. This external validation is looked upon as step by which internal bias is minimized, the rigor of the evaluation can be controlled, and accountability maintained. Johnson and Christal (1985) believe that "although accreditation is a useful process that can promote self-analysis, it remains both periodic and externally motivated. Self-study, on the other hand, can become an integral and continuing part of an institution's self-consciousness" (p. 1). While this statement is helpful in determining the difference between accreditation and self-study, it also places an emphasis on maintaining the process of evaluation as an ongoing part of the institution's functioning.

Self-study is a term which represents the concept of institutional self-evaluation. The actual process of self-study varies upon the implementation and can either be self-initiated or mandated. In either case, most documented uses of self-study include a portion of the process which involves the staff comparing the school program to a set of indicators of effective programs or quality indicators. The indicators of effective programs are usually produced by curriculum or subject specialists external to the school. The self-study concept can embody many great processes for stimulating introspection on a school wide basis, but the use of predetermined curriculum based indicators in judging school programs introduces the bias of the subject specialist and may not allow for the discovery of unanticipated program outcomes. The direct comparison of a program with set standards, creates a checklist mentality, stifling creativity and uniqueness. The only advantages that checklists have is in helping schools focus their evaluations, and in the time which is conserved. Other aspects of self-study involve ongoing discussions amongst the staff, the collection and analysis of data and report writing.

The advantage represented in the processes set out by the California State Dept. (1987a, 1987b), is that the process requires observations, the collection of data and collaborative discussion prior to arriving at a basis for comparison. A common perspective amongst the group must be derived prior to using the quality criteria set down by the State. Areas in which the program match or exceed the description are considered strengths, while others which do not meet the stated criteria, are considered areas for improvement. The group then must decide which of the areas for improvement should gain the most priority. Lastly, after sequence is considered, an action plan is created for improvement. The conclusions of the group are recorded in a format which will facilitate inclusion in the total school summary, then shared with the rest of the school, and eventually with the review committee.

(1987a, p. 54-55) This practice reinforces the team concept of working together to accomplish a goal, rather than working in isolation. The process also allows the school to use evaluation as an ongoing part of its' decision making process.

The work of Clift, Nuttall and McCormick (1987) provides a reasonable overview of the research into school self-evaluation in Britain. It is their collective belief that "the fundamental purpose of school self-evaluation is the improvement of the quality of education" (p. 190). They cite four basic strands of development for which self-evaluation has impact: accountability, curriculum development, program review and staff and institutional development. Although they state that "accountability and school improvement are inextricably linked" (p. 190), it is their belief that outside pressure on schools to evaluate, might be detrimental to "the development of a genuinely open spirit of self-evaluation" (p.190). Therefore their work is primarily based on school initiated self-evaluation.

One approach used in Britain is termed the GRIDS project (Clift, Nuttall and McCormick, 1987, p. 6). It is a voluntary procedure for the self-evaluation of schools that is premised on the involvement of the entire staff. The second most prominent principle which this evaluation incorporates is the importance of keeping accountability considerations out of the evaluation process. As with most of the British program reviews, this one involves the analysis or development of statements of aims or objectives. Clift et al. claim "this collegial/democratic model of evaluation is seen as the most likely to promote an honest and thorough evaluation with the best chance of change being implemented" (p. 6). It is unfortunate that their work did not provide a more detailed description of the project.

The Solihull evaluation scheme was one of the first models actually published in Britain, for schools to use. It contained 41 pages of statements, grouped under four main headings; "evaluating organization and management, evaluating specific aspects of the school's work, further analysis of teachers' role [and] inservice education and staff development" (Clift, Nuttall and McCormick, 1987, p. 55-56). Each category contained a series of subheadings with a plethora of statements following. The concept was well intended, but unfortunately, no set direction was provided for utilization of the booklet. Clift et al. provided no indication of what the statements were based upon, but did indicate that teachers had to react to the statements. Those schools who chose to participate used it in drastically different ways. The results indicated that directions for use are important, as is the significance which the administration places on the activity. In general, the staff found the process of guided introspection an effective means of stimulating professional development.

Activity evaluation is the name given to internal evaluations that are performed at universities and colleges in Sweden. The aim is to provide a better basis for decisions internal to the institution. Adelman and Alexander (1982) describe the process as paralleling institutional evaluation in the United States. The process is very highly planned and structured. The procedure simply stated, first involves the description of the program in as complete a manner as possible. Then a great deal of time is spent in group discussion of the goals, objectives and philosophy of the program. Then the data is collected and analyzed, and the process finishes with discussion centered around improvement. Reports are generated in a format which assists in the utilization of the information in the improvement process.

In Canada, school self evaluation procedures differ widely. Alberta teachers began their self-evaluation of schools in 1969 with a publication produced by the Alberta Teachers' Association. It was created with the intent that "elementary teachers will accept the premiss that significant improvement can result from a sincere and objective staff examination of the school situation" (Alberta Teachers' Association, 1969, p. iv). It was hoped that

> the guide will serve to stimulate staff discussions regarding effective practices, focus attention upon desirable standards, and provide a broader and clearer concept of the character and function of the effective elementary school. In general, the ultimate purposes of self-evaluation are:
>
> a) clarification of educational aims and objectives,
> b) improvement of instructional programs,
> c) increased competence of personnel,
> d) better utilization of facilities, and
> e) improved coordination of total staff effort. (p. vi)

The procedure also provided a proposed time line, with a built in option of external evaluation upon completion of the self-study (accreditation style). The package provided many prepared forms for data collection. The methods of evaluation centered around rating the school's performance on predetermined criteria. The stated purposes for the publication are very similar to the rationales presented for other models of internal or self-evaluation. This procedure was rarely used by schools. The reasons for this can only be surmised, as documentation is unavailable. The most probable causes: the educational climate of the day was not strongly oriented towards the search for excellence; and, management of schools was popular rather than leadership through democratic means. The possibility exists that schools failed to find the process useful.

This manual was updated and republished in 1989 under the title *School and Program Evaluation: A manual for Teachers*. The format changed somewhat, but

the goal of the process was similar. The major change was the broadening and rephrasing of the aims.

> In general, the aims of school and program evaluation are:
> a) clarification of educational aims and objectives,
> b) identification of strengths,
> c) improvement of instructional programs,
> d) professional growth of teachers,
> e) better utilization of facilities,
> f) identification of needed improvements to facilities, and
> g) improved coordination of total staff effort. (p. 3)

A similar document for the evaluation of elementary schools has been prepared in British Columbia by the West Vancouver School District #45 (1986). This booklet was patterned after an accreditation scheme prepared by the British Columbia Department of Education for high schools. The booklet is to be used in work book fashion, where answers and information are recorded directly. Again the staff responds to a series of predetermined questions, prepared to guide the collection of data. This system is slightly more useful than the one from the Alberta Teachers' Association, in that its questions remain open ended, allowing for thought, discussion and reaction rather than direct rating. This process also calls for an external evaluation upon conclusion of the self-study. It is recommended that the procedure be undertaken once every five years.

The evaluation publication produced for the Province of Saskatchewan is an outcome of its' educational review. Burgess (1987), the editor of the manual, states that the Saskatchewan School Improvement Program is designed "to build the capacity for improvement at the school level in order that Saskatchewan schools may put into action the best practices to ensure student learning and success"(p. 1.1). This program of school improvement has eight essential elements, of which one is "systematic monitoring and evaluation" (p.1.1-1.2) Burgess further elaborates the

link between effective schools research and the program for monitoring and evaluation.

> effective schools have systematic procedures and strategies to monitor and evaluate priorities, programs, students and teachers. In such schools information is collected in a variety of ways on an ongoing basis, and is then used in decision-making and goal-setting for improvement". (p. 1.2)

Burgess indicates that although educators have always made decisions, this new era brings with it "an orientation to program evaluation as the normal professional stance" for program change and improvement (P.1.2). It appears clear then that schools in the Province of Saskatchewan have been directed to make use of school-based program evaluation, as part of their school improvement strategies.

The manual which has been produced to guide the process, is a comprehensive package, which appears to have real potential for school use. It is divided into six sections comprised of introduction, planning and doing an evaluation, curriculum areas and programs, evaluation approaches, data collection, and evaluation models. When combined, the sections provide school personnel with the theory behind quality evaluation, a step-wise guide outlining strategies for planning and conducting an evaluation, specific concrete examples for illustration, discussions of school programs with sample evaluation questions, a variety of suggestions for data collection complete with examples and a choice of evaluation models. The manual provides the user with an overview of the major models in program evaluation, a statement of appropriate use for each, and bibliographic reference for further reading on that model.

This approach which has been prepared for Saskatchewan schools, appears to have a great deal of promise, but the true test will be in the extent to which it is used. From the statements made by Burgess (1987), the overall strategy has the blessing of

the Department of Education, as an expectation of normal professional activity. Common sense suggests that this may help to overcome the problem of emphasis upon implementation which engulfed the Solihull experience in Britain. Because the emphasis of this approach to program evaluation is program improvement and not accountability, its attempt to incorporate more traditional models of program evaluation into a self-evaluation process should be praised.

Another advantage of the Saskatchewan publication is the way it points out that evaluation can and should be used to enhance the individual programs which the school offers. This is a step forward from the Alberta and British Columbia models which deal only with the school as a whole. Further more, the manual demonstrates ways in which curricular resources, such things as text books and computer software can be evaluated as part of the individual program evaluation.

Problems and Promises

One of the major problems facing program evaluation is the utilization of program evaluation information by schools. To help understand this problem, several researchers have done work on utilization of evaluation results, and provide us with some interesting insights. One of the prominent researchers into utilization of evaluative information is Marvin Alkin. Many researchers have used the foundation which he and his co-workers provided. Alkin et al. (1979) presented two views of evaluation usage: the first that evaluation is not used to make decisions, and that there is little hope that it ever will; and the second, stating that decision makers already use evaluative information, but in an indirect fashion which is not always visible. In later work, Alkin and Stecher (1983) present the idea that decision making is a complex task. Their study indicates that in fact, decision makers utilize

information from one type of evaluation, needs assessment, in what they refer to as the recognition phase of decision making.

> Needs assessment helped school staff identify areas that required attention. By evaluating the status of the school program on an annual basis in a form that was familiar and in a manner that involved the staff directly, needs assessment had a sizable impact. (Alkin and Stecher, 1983, p.29)

On the basis of their results, Alkin and Stecher point out the need for information to be generated locally for it to be utilized by the school. They also point out that the personal investment which the staff have in generating the information, has an effect on its utilization. Their study provides a number of good arguments for a movement of program evaluation into the schools. Their research provides ammunition for Stake's responsive model of evaluation, when they claim that for "evaluation efforts to contribute to school improvement they must be motivated out of local concerns and must serve local needs" (p. 30). They conclude by advocating that "the capability for performing evaluation must be shifted to the local schools themselves" (p.30).

Love (1983) in his assessment of internal evaluations stated that the major problems with internal evaluation stem from dilemmas posed by the context of the organization. Love contends that the use of program evaluation for program improvement purposes "assumes that managers and professional staff want to improve the performance or their programs" (p.7). Most school districts have ongoing program improvement as an expectation of normal school functioning. Therefore, given the present educational climate, this is a valid assumption for the majority of schools. Love also indicates that the presence of an evaluation component influences human behavior and that "the behavioral consequences of evaluation should be viewed as an integral component of the evaluation process" (p.

8). He further suggests that this behavioral influence will effect decision making. Love believes that the structure of the organization is one of the greatest factors to affect the implementation of internal evaluation. Love states that "specific types of organizations appear to possess psychological climates that nurture the evolution of self-evaluation" (p. 14). There is a link between this climate and the type of organization which can successfully use internal evaluation. Love indicates that the ideal type of organization is called "participative". In the following paragraph, Love describes the organizational factors which bind the organization to its management system.

> These factors include philosophies, values, standards, methods of decision making, lines of authority and responsibility, and use of resources. The policies, values, and philosophies of the participative organization encourage supportive behavior throughout the organization. Information is shared, and it is used for problem solving throughout the organization. Workers are trusted. Clear communication is actively encouraged. High standards of quality are maintained through regular evaluation of results. Loyalty to the work group and the organization is fostered. The contributions of workers and management are valued. (1983. p.15)

This description of organizational climate points out the complexities involved in making internal evaluation work. It reiterates the basic concepts of educational philosophy and autonomy upon which Stake built his responsive evaluation model, and which are needed for education to move towards excellence. Love also suggests that internal evaluation provides support for strategic planning, by collecting valid data upon which to base decision alternatives.

Not all schools possess the participative climate or structure which is required for internal evaluation to be successful. The climates which exist in schools are largely attributable to the institutional leader, the principal. Work by Newman, Brown and Rivers (1983) indicate that a psychological construct called "locus of control" is linked to the need for, and utilization of evaluative information.

Specifically, they found that persons who perceive themselves as "internals" or having control over their own fate, tend to see the need for evaluation, welcome information from significant others, and tend to find group decision making useful. It appears that locus of control for administration may be beneficial as one measure in predetermining the success of a participative climate in the school and subsequently, internal evaluation.

A participative climate in a school will include the use of the teaming concept for internal evaluation. This method when used in industry is called participative work groups. Vogt and Hunt (1988) provide some valuable insights into work groups, and why they fail to last in American business. These insights have direct implications for attempts to incorporate internal evaluation into schools.

The primary function of the work group is to identify, analyze and solve problems in their work area. Although many groups are successful in solving problems, their main downfall appears to be their general lack of support from management. Vogt and Hunt elaborate on several of the factors which cause this deterioration. Work groups require a participative culture in the organization for to be successful. This is partly because the work group concept is based on active employee involvement in decision making. These decisions are evolved using consensus. Learning that they can have a voice in decision making, eventually increases workers desire to influence policy decisions. Management is not always pleased or prepared for this eventuality.

Vogt and Hunt state that "effective training is key to the success of work groups. Critical problem areas range from not learning the specific guidelines for brainstorming to not understanding the role of healthy conflict" (Vogt & Hunt, 1988, p. 100). Often, work group facilitators have too many other responsibilities which

take priority over the work group. In addition, they observed that work groups do not work well in climates of uncertainty including employee turnover, layoffs, and union conflicts. For this methodology to flourish, management must show its support for participative strategies, by both providing the time and human resources necessary for task accomplishment. This summary of participative experiences in industry provides great insight into the organizational structure a school must possess, if it is to make use of the self-evaluation concept.

Clift, Nuttall and McCormick, (1987) express several points in conclusion to their six years of study on school self-evaluation which sheds a cautiously pessimistic light on its future. These studies dealt with total school evaluation in all but one case, and the methods varied from checklist through departmental discussion, to a form of accreditation. Despite the variety of methods, and the case study approach to this research, several of their comments are valuable.

> In its idealized form, school self-evaluation requires idealized schools in which collegiality, cooperation, open communication and fraternity rule and where professional development and professional self-respect go hand in hand. A measure of egalitarianism is probably vital. Not many schools match this in reality! (Clift, Nuttall and McCormick, 1987, p. 209)

Much of what is contained in the concept of "idealized school" can be interpreted as parallel to the participative climate espoused by Love (1983). It is arguable just how many schools meet these standards at present, but if in schools search for excellence, internal program evaluation is to take hold, it will be important to know the conditions which will foster its growth.

> School self evaluation requires a highly professional teaching force, trained in the skills of institutional review, aware of and confident in their own professionality, possessed of the high morale necessary to seek for constant improvement in the quality of education which they offer and confident of support in this from the other stakeholders in education. The institutionalization of SSE [school self-evaluation] requires a different

> apportionment of teachers' time and energy than that pertaining in most schools at the present time. (Clift, Nuttall and McCormick, 1987, p. 210)

Again the need for staff training becomes evident. The professional growth of teachers is a continual processes which takes place slowly over time. As was mentioned by Love (1983), in order for improvement to take place, both teachers and administration must want it to happen. For teachers to want this to happen, surely professionalism must run deep enough. The time factor is something which will always be a problem in education. Teachers are continually asked to take on added responsibilities, in the name of what is good for kids, without added time, energy or monetary compensation. Vogt and Hunt (1988) stated that if management was intent on the work group process succeeding, then they must be willing to provide the resources, both time and human, to get the job done. A collaborative effort by teachers and administration will find the time if they feel that the process is truly beneficial, and will move them closer to excellence.

Summary

The review of the literature has found an expressed need for school-based program evaluation, and provided a rational for its development. It should be regarded as a modification of existing program evaluation methodology (responsive evaluation), intent on the improvement of educational programs. Program evaluation is already utilized in a very informal manner in most schools as needs assessment. If schools are serious about moving from competence towards excellence, a comprehensive and ongoing process of introspection will be required. The following points will outline the important concepts of such a process, and provide the groundwork for further development in school self-evaluation.

The evaluation should:

1. be internally controlled.

2. be improvement oriented, not intentionally geared for accountability or advocacy.

3. be responsive to the concerns and needs of the school and community. (as in Stake's responsive model)

4. provide a detailed description of the program as it was intended to operate.

5. reassess the philosophy and goals upon which the program was based, through meaningful discussion.

6. function as a management tool, providing information for decision making.

7. be group directed. The group to be composed of interested stakeholders, and coordinated by an administrator.

8. utilize the program staff and other stakeholders as much as possible in the process. (describing, focusing, designing, collecting, analyzing, comparing, reporting, prioritizing and planning for improvement.)

9. be focused on the basis of information needs.

10. design data collection to utilize as many methods and sources as possible.

The following factors have been shown to influence the success of evaluations.

1. Both the staff and administration must possess a strong desire for the improvement of education in their school. This desire must be strong enough to welcome change in both school structure and approach to educational leadership. (If this is true, it is likely that the principal perceives that he/she has the potential to improve the quality of education.)

2. The school must possess or work towards developing, a participative climate of leadership in which:

- shared decision making is the norm throughout the organization.

- philosophies, values, standards and policies are derived through consensus, to build a common basic language with which to describe their practice.

- staff are trusted.

- loyalty to the school is fostered.

- information is shared throughout the organization, to assist decision making.

- the contributions of individuals are truly valued, not just expected.

Conclusion

It becomes apparent from the literature, that for formal evaluation to become utilized at the school level, will require far more than just another model of evaluation. Without utilization of the information provided in an internal evaluation, the benifits gained by involvement in the process may go unnoticed. The influence of organizational factors on the implementation and utilization of internal evaluation are far greater than anticipated. For schools to participate in an internal evaluation, they must choose to do so, based on the firm belief that it will be of direct benefit to them. Internal evaluation for any other reason will result in an evaluation of less than acceptable standards.

For schools to participate in internal evaluation, they will need a resource manual similar in design to the one produced by Burgess (1987), for use in Saskatchewan. This will provide an outline of the various models of evaluation, as well as useful guidance for planning and carrying out the evaluation. Help is provided for all steps along the process continuum, until the evaluation reaches its conclusion.

Further research is needed into the utility of such a procedure for small scale, or single topic evaluations in schools. An example of such research could involve evaluating the computer program in schools, in varying stages of participatory organizational structure, and determining the extent to which staffs perceptions differ.

Further research is needed into the attitudes of school-based staff toward participation in program evaluation. In particular, insights are needed which illuminate the perceived benifits and drawbacks experienced during involvement in program evaluation. The benifits and drawbacks should be structured to reflect the important aspects of internal evaluation which were exposed in this review of the literature.

Chapter Three

Design and Methodology

This chapter contains the details of methodology and procedure for sampling, data collection, instrument design and data handling. This includes a complete description of the screening, analysis and interpretation of the data as it applies to each subproblem.

The intent of this research is primarily descriptive, requiring the collection of data as observed. In this study, observations were collected using a survey method.

Population and Sampling

The sample used in this study was drawn from within the Calgary Board of Education. The population includes all teachers and principals in elementary, junior high and senior high schools, but excluding those in schools which could be classified as a mixture of these school types. This is important so that accurate comparisons of attitudes can be made between school types. Due to financial and time constraints, a sample limit of 200 was used. The actual sample size was 199 persons not the expected 200. This was arrived at by creating a proportional stratified random sample from the population of subjects, classified by school type. A sample of this type was used to provide an accurate representation of the population. The sample was randomly created using a computer and the most recent personnel data files of the Calgary Board of Education.

Confidentiality

Because the survey dealt with attitudes of staff towards a sensitive topic, the identity of the respondents was protected to ensure the highest possible rate of

return. The respondents received a numbered envelope in which to return the survey. The numbers were used to follow up non-respondents, while assuring anonymity. Respondents were not required to place their name on the survey. Records of the name-number match were strictly protected to assure security for the respondents.

Data Collection Techniques

The instrument was distributed to the participants by June 1st, 1990 and by July 9th, 97 of the 199 had been received. The inter-school mail system was the vehicle for the distribution and return of the surveys. Permission for the use of this system was obtained from both the Department of Research and Testing and Mr. Don Royan, Assistant Area Superintendent. The respondents were provided with an addressed return envelope in order to make the process as easy as possible for them.

Participants were given two weeks to return the questionnaire prior to the first follow-up. After two weeks, a phone call was made requesting return of the questionnaire. Those who needed another copy of the survey were sent one. During the last week of June, a final follow-up was conducted by telephone. Ten non-respondents were contacted at random inquiring about reasons for non-response, and collecting demographic data to assure that non-respondents were not all from a particular sub-group within the population. To assure that the non-respondents views did not differ substantially from those of the respondents, a random telephone survey was conducted. Three non-respondents from each school-type were chosen at random and asked to complete the instrument over the phone. The results were then compiled and compared with those of the respondents. No substantial

difference was shown to exist. It can then be concluded that the non-respondents do not represent any specific sub-group within the sample or population.

Research Instrument

The instrument used to collect data included carefully constructed attitude statements which Turner and Clift (1985) used in their Solihull study. They gathered these statements over time from teachers' comments about program evaluation. The extent of these comments included comparing internal evaluation with external evaluation, questioning the capacity of program evaluation "for probing institutional processes in depth, and whether it motivated change" (p.105). Some statements were deleted because they were designed to collect attitudes strictly based on formal-internal evaluation. A few statements were added to reflect issues which other authors had determined to be significant and were directly applicable to the Indicators of effective programs produced by the Calgary Board of Education (1989). Lastly, some British terms were modified with language more appropriate to Calgary.

The instrument was pilot tested to establish face validity and reliability with a group of 12 teachers not involved in classroom teaching. Changes required as a result of this testing were incorporated in the instrument prior to data collection.

The survey was constructed in four parts. The first gathered demographic data to help in the creation of subgroups within the population. The second assessed the experience the staff had with evaluating programs, in either formal or informal terms. The third section examined the extent to which staff agreed with the attitude statements provided about program evaluation. The last section determined general impressions about the evaluation of programs without respect to category or type. These last three used a Likert-type 5 point attitude rating scale ranging from strongly agree to strongly disagree, with uncertain as the midpoint.

Five categories of attitudes relating to program evaluation were derived by Turner and Clift (1985). These are collegiality and openness, professionality, efficiency and threat, professional development, and formality. Statements were created to assess the attitudes of teachers in each category. To arrive at a common meaning of these terms in Calgary, they were pilot-tested by teachers and consultants who had participated in program evaluation. The results have been compiled and are discussed with the results of subproblem four in the next chapter. Participants in this pre-study were excluded from the population sampling process. Using this methodology avoided contamination of the sample in any way.

The Data

The data are of two types, primary and secondary. The primary data were collected through observation. The secondary data were published results from the Solihull study conducted by Turner and Clift (1985) which found five basic categories of attitudes toward program evaluation. Further discussion of the data is provided in the context of each subproblem.

For subproblems one through four, the primary data required were collected from the returned surveys.

Subproblem 1

To what extent have staff been involved in "evaluating" school programs, either formally or informally?

a) Determine what proportion of staff have been involved in the evaluation of programs.

b) Categorize the type of evaluation staff have participated in as either informal or formal with internal or external control.

c) Determine if similarities or differences by subgroup exist.

The data needed. The data needed for solving subproblem one were a) the responses to the survey observing the actual number of participations in program evaluation activity, b) the responses to the survey observing the actual number of participations in program evaluation activity categorized by type, and c) responses to the survey providing demographic data to determine subgroups.

How the data were screened. The responses of the participants were screened to assure that only data was used from respondents who completed the survey according to the directions given.

Analysis of data. The data were collected in categories which represented increments greater than one. Therefore the data is of the ordinal variety and was analyzed using descriptive statistics. The responses to the question of evaluation participation were compiled using an item analysis to determine the general extent. The resultant numbers were converted to a per cent value. The data values were then grouped and compared according to subgroups of teaching experience, gender and school type. The results were also compared across evaluation types, to determine if differences existed.

Interpretation of the data. Means of the data were compared by classification of evaluation. The major goal of this section was to describe the extent of evaluation practice within the Calgary Board of Education. Trends and patterns which existed were elaborated.

Subproblem 2

What are the perceived benefits and drawbacks of program evaluation?

a) Evaluate the views of all respondents, as to the perceived benefits and drawbacks of the evaluation types which they have experienced.

b) Determine the differences which exist between different categories of program evaluation.

c) Assess the mean attitudes of staff within the following groups which make up the population: school type, gender groups and experience groups.

d) Determine if there is a difference in attitude among the subgroups.

The data needed. The data needed for solving subproblem two were a) the responses to the survey reporting the extent of agreement or disagreement with twenty concepts associated with the evaluation of programs grouped by the three basic evaluation classifications, and b) responses to the survey providing demographic data to determine subgroups.

The means of obtaining the data. The research instrument was constructed and distributed as outlined in the sections "Population and Sampling, and "Research Instrument". The data were gathered through the use of these specific procedures.

How the data were screened. The responses of the participants were screened to assure that only data was used from those who stated having experience with that specific type of evaluation. This assured that attitudes and opinions expressed about specific evaluation types, reflected an experiential basis. Responses that did not

reflect an experiential base were excluded from the analysis. Responses which reflected more than one answer for a single statement were omitted.

Analysis of data. The data were analyzed through the use of descriptive statistics including means, frequencies and per cents. Analysis included the crosstabulation of data by type of evaluation and by subgroup within that type of evaluation. Several tests of significance were used. Chi square was used to determine the significance of frequency distributions. Analysis of variance was used to determine the extent to which differences among groups occured beyond chance expectation.

Interpretation of the data. The task of descriptive research is to present a clear picture of the situation as it appeared at the time of data collection. As such, tables and text are used to describe the data, including patterns or trends which appeared. Mean values are used to interpret the extent to which groups agree or disagree with the statements. Values between 3.1 and 5 reflect agreement, with the degree of agreement increasing with the value. Values between 2.9 and 1 represent disagreement, the strongest of which could be 1. A value of 3 signifies a group or collective indecision.

Differences among group means will be reported if they exceed .4. Results are considered significant at the .05 level.

Subproblem 3

Determine the views of all respondents on selected issues relating to the benefits and drawbacks of implementing program evaluation.

a) Assess the mean attitudes of staff as a whole for each issue.

b) Determine the mean attitudes for each issue, grouped by school type, gender and experience.

c) Determine if there is a difference in attitude among the subgroups for each issue.

The data needed. The data needed for solving subproblem three included the responses to the survey reporting the extent of agreement or disagreement with seven "general impressions" statements about program evaluation. The "general impressions" section did not require specific experience.

How the data were screened. Responses which reflected more than one answer for a single statement were omitted.

Analysis of data. The data were analyzed through the use of descriptive statistics including means, frequencies and per cents. Analysis included the crosstabulation of data by type of evaluation and by subgroup within that type of evaluation. Several tests of significance were used. Chi square was used to determine the significance of frequency distributions. Analysis of variance was used to determine the extent to which differences among groups occured beyond chance expectation.

Interpretation of the data. The task of descriptive research is to present a clear picture of the situation as it appeared at the time of data collection. As such, tables and text are used to describe the data, including patterns or trends which occur. Mean values are used to interpret the extent to which groups agree or disagree with the statements. Values between 3.1 and 5 reflect agreement, with the degree of agreement increasing with the value. Values between 2.9 and 1 represent

disagreement, the strongest of which could be 1. A value of 3 signifies a group or collective indecision.

Differences among group means will be reported if they exceed .4. Results were considered significant at the .05 level.

Subproblem 4

To what extent do the attitudes of staff fall into the categories determined to exist by Turner and Clift (1985) in a similar study of schools in Solihull England?

a) Duplicate the process of factor analysis undertaken by Turner and Clift (1985).

b) Make corrections for terminology between England and Canada.

c) Compare and contrast the results.

The data needed. The data needed for solving subproblem four included the responses to the survey reporting the extent of agreement or disagreement with twenty attitude statements about program evaluation. Specifically, the data used by Turner and Clift (1985) involved internal evaluation. Two types of internal evaluation were included, both informal and formal-internal evaluation. Therefore only the data collected which reflects this form of evaluation is used in the comparison.

How the data were screened. The responses of the participants were screened to assure that only data were used from those who stated having experience with that specific type of evaluation. This assured consistency in the comparison of these two studies. Responses which reflected more than one answer for a single statement were omitted.

Analysis of data. The data were analyzed using the same procedures as Turner and Clift (1985). This included the use of factor analysis, with the resulting factors being rotated to an orthogonal solution (Varimax). This rotation of the factors was done to reduce the ambiguity of the factor loadings, and make meaning become more evident. Two methods were used in the exploratory factor analysis; Principle component analysis (common factor method) and Iterative principle axis factor analysis. No difference in results was produced.

Variations in factor analysis occur because of the methods used in the research. One such variation is caused because the the correlation matrix upon which the analysis is based, is created from sample data rather than the entire population. "For any given sample, an observed correlation will never exactly reflect the underlying population correlation." (Kim & Mueller, 1978a, p. 64) A number of different methods must be used to provide the researcher with various possibilities, but the final decision remains a rationalization of the most logical solution.

Another possible source of error in factor analysis lies in the selection of variables to examine. The variables dealt with in this research, are just a small subset of all the possible variables.

> In general, the greater the ratio of the number of variables to the number of underlying factors, the more informative the factor analsis is. But what is crucial is not the overall ratio, but the number of variables for each factor. (Kim & Mueller, 1978a, p. 68)

Kim, Mueller and others prefer to see at least three variables per factor in order to provide reasonable results. In additon, most methods used in factor analysis assume that there has been no sampling of variables.

Interpretation of the data. The factors were interpreted by reference to statements with loadings of (0.4) or greater. Determination of actual factors involved a rationalization of the most logical solution given the alternatives presented. Interpretation of the factors was based on two basic assumptions, the "Postulate of Parsimony" and the "Postulate of Factorial Causation".

Postulate of Parsimony: Given a number of equally compatible models, only the model with the least number of common factors is believed to be true. (Kim and Mueller, 1978b, p.86)

Postulate of Factorial Causation: It is assumed that the observed variables are linear combinations of underlying factors, and that the covariation between variables is due to their common sharing of one or more common factors. (Kim and Mueller, 1978b, p.86)

The Scree-test method designed by Cattell (1965) was used to determine the maximum number of common factors. "Some Monte Carlo studies indicate that his method is often superior to others where there are minor factors and the interest is in locating only the major common factors." (Kim and Mueller, 1978b, p.45) The eigenvalues test was also applied as another method of determining the number of factors.

Chapter Four

Presentation and Analysis of Results

This chapter presents the data and findings in a structure which uses the subproblems as an outline. The findings are preceded by a discussion of related demographic data.

Sample Demographics

The instrument was distributed to participants using a proportional random sample technique. Table 1 illustrates the distribution of the participants proportionally by school type.

Table 1

Distribution of Participants

	Sample	
School Type	Frequency	%
Elementary	111	57.8
Junior High	39	15.4
Senior High	49	26.8
Total	199	100.0

Of the 199 instruments distributed, 97 were returned. This equates to a return of 48.8% of instruments. Elementary and senior high schools both exceed a 50% return while the junior high sector showed the lowest response rate at only 38.5%.

Table 2 displays the distribution of respondents and the percentage of return for each school type.

Table 2

Distribution of Respondents by School Type

	Respondents	
School Type	Frequency	%
Elementary	56	50.5
Junior High	15	38.5
Senior High	26	53.1
Total	97	48.8

Several subgroups were initially identified to be of interest in this study. Information was then collected to assist with the formation of these subgroups within the sample. One subgroup determined to be potentially important, was the type of school in which the teachers taught. Table 2 illustrates the distribution of the respondents by school type.

A second subgroup selected was years of experience in education. Experience was responded to in three categories: 1 to 5 years, 6 to 10 years and 11 or more years. The results show 16.5 % of the sample with 1 to 5 years, 20.6% with 6 to 10 years and 62.9% with 11 or more years. It may be generalized that the majority of the respondents were well experienced educators.

Gender was the third subgroup of interest. Females represented 69.8% of the sample. This equated to a proportion of 2.3 females for every male in this sample. In elementary school, there was a 5.2 female to male ratio. In high school, the ratio was reversed with a slight male to female advantage of 1.27 to 1.

Extent of Involvement in Program Evaluation

Subproblem one focused on determining the extent of staff involvement in program evaluation, and to classify it by the formality and locus of control.

Data were collected concerning the extent of participation in program evaluation using four different questions for purposes of classification. In the demographic section of the survey, the respondents recorded the extent of their involvement in the evaluation of a school-based program. This category of evaluation was called "General". Three other subclassifications made distinctions between the formality of the procedure and the locus of control for the evaluations. "Informal" evaluation specified that the control of the evaluation remained with the teacher, and no formalized procedure was followed. Although the label "Informal" does not explicitly state the locus of control, internal is implied. The last two subclassifications, "Formal-internal" and "Formal-external", required set or accepted procedures of evaluation but varied the control. Internal being teacher or school controlled as contrasted with external which was directed by a person outside the school.

Table 3 reveals the overall picture of staff involvement in program evaluation and compares the three subclassifications of program evaluation.

Table 3

Extent of Involvement in Program Evaluation by Classification

Results expressed as per cent.

	Type of Program Evaluation			
	General	Informal	Formal - Internal	Formal - External
n	97	95	90	81
Number of Evaluations				
0	6.2	16.8	35.6	58.0
1-2	37.1	30.5	46.7	32.1
3-5	28.9	25.3	11.1	6.2
6-10	12.4	11.6	4.4	2.5
11+	15.4	15.8	2.2	1.2
Total	100	100	100	100

Only 6.2% of the respondents indicated having never participated in any program evaluation. Table 3 shows the majority of respondents involved in informal program evaluation to have participated between one and five times. In the formal categories, the majority participation ranges from zero to two times. This reveals a general trend to less involvement in program evaluation, as the process becomes formal and the control changes to external. The majority of respondents had never participated in a formal-external evaluation while involved in education.

Table 4 shows no significant relationship between years of experience and general involvement in program evaluation. Of interest is the observation that 65.9% of the sample had experienced some form of evaluation only one to five times, when 83.5% of the sample had been teaching for more than five years. Of the 61 respondents who were classified as experienced teachers (11+ years) only 13 had evaluated programs 11 or more times. Of the 20 teachers with 6 to 10 years experience, 9 had only one or two experiences with evaluation. It is apparent that teachers in this sample had reasonable experience in the general evaluation of school-based programs.

Table 4

Distribution of Involvement in General Program Evaluation by Years of Teaching Experience (expressed in per cent).

	Teaching Experience (Years)			
Program Evaluations	1-5	6-10	11+	Total
n	16	20	61	
0	1.0	1.0	4.1	6.2
1-2	10.3	9.3	17.5	37.1
3-5	4.1	7.2	17.5	28.8
6-10	0.0	2.1	10.3	12.4
11+	1.0	1.0	13.4	15.5
Total	16.4	20.6	62.8	100

Experience with informal evaluation appears just as varied as it was for general evaluation as described in Table 5. Of those teachers with 11 years or more experience, 32% had only experienced informal evaluation once or twice while

another 32.2% had experienced this type of evaluation in excess of 5 times. Of the teachers with 6 to 10 years experience, 65% had experienced informal evaluation 1 to 5 times. The majority (56.2 %) of those teachers in their formative years (1-5 years), reported informal evaluation experience 1 to 5 times, while a surprising 25% indicated experience in excess of 5 times.

Table 5

Distribution of Involvement in Informal Program Evaluation by Years of Teaching Experience (expressed as per cent).

	Teaching Experience (Years)			
Program Evaluations	1 - 5	6 - 10	11+	Total
n	16	20	59	
0	3.1	4.2	9.5	16.8
1-2	5.3	5.3	20.0	30.6
3-5	4.2	8.4	12.6	25.2
6-10	2.1	1.1	8.4	11.6
11+	2.1	2.1	11.6	15.8
Total	16.8	21.1	62.1	100.0

Examination of informal evaluation experience by school type shows an interesting but not statistically significant difference between elementary school teachers responses and those of junior or senior high teachers. The major distinction is in the different proportion reporting no experience of this type. Only 10.7% of elementary teachers reported no informal evaluation experience, while 28.5% was reported by junior high and 24% by senior high teachers.

Analysis of formal-internal and formal-external evaluation experience by subgroup provided no significant findings. The number of respondents indicating involvement in formal-internal program evaluation once or twice, was distributed in approximate proportion to both experience and school type of the subgroups. This is demonstrated by tables 6 and 7.

Table 6

Extent of Involvement in Formal-Internal Program Evaluation by Teaching Experience.

Results are by per cent of column.

	Teaching Experience (Years)			
Program Evaluations	1-5	6-10	11+	Total
n	14	20	55	
0	35.7	35.0	36.4	36.0
1-2	50.0	45.0	45.5	46.1
3-5	14.3	10.0	10.9	11.2
6-10	0.0	5.0	5.4	4.5
11+	0.0	5.0	1.8	2.2
Total	100.0	100.0	100.0	100.0

Table 8 illustrates that the per cent of each experience group which has participated in formal-external evaluation, increases with years of experience.

The analysis of formal-external evaluation indicates that teachers at the high school level have participated to a greater extent than teachers at the other levels.

Table 9 also illustrates that elementary teachers have a higher percentage participation than their junior high colleagues in this type of evaluation.

Table 7

Extent of Involvement in Formal-Internal Program Evaluation by School Type

Results displayed as per cent column.

	School Type			
Program Evaluations	Elementary	Junior High	Senior High	Total
n	50	13	26	
0	36.0	23.1	42.3	36.0
1-2	40.0	61.5	50.0	46.1
3-5	16.0	15.4	0.0	11.2
6-10	4.0	0.0	7.7	4.5
11+	4.0	0.0	0.0	2.2
Total	100.0	100.0	100.0	100.0

Benefits and Drawbacks

Subproblem two sought to determine the perceived benefits and drawbacks of program evaluation. The results are discussed in the context of the three basic classifications which were chosen as important in the utilization of evaluation information. Within the three classifications, subgroup differences will also be discussed. In order to respond to the attitude scales, participants were required to have experience with the specific classification of evaluation.

Table 8

Extent of Involvement in Formal-External Program Evaluation by Teaching Experience

	Teaching Experience (Years)			
Program Evaluations	1-5	6-10	11+	Total
n	4	7	23	
0	69.2	63.1	53.1	58
1-2	30.8	26.3	34.7	32.1
3-5	0.0	5.3	8.2	6.2
6-10	0.0	5.3	2.0	2.5
11+	0.0	0.0	2.0	1.2
Total	100	100	100	100

Table 9

Extent of Involvement in Formal-External Program Evaluation by School Type.

Results are expressed as a per cent of column.

	School Type			
Program Evaluations	Elementary	Junior High	Senior High	Total
n	18	3	13	
0	59.1	75.0	48.0	58.0
1-2	34.1	16.7	36.0	32.1
3-5	0.0	8.3	16.0	6.2
6-10	4.5	0.0	0.0	2.5
11+	2.3	0.0	0.0	1.2
Total	100	100	100	100

The rating items from 1 to 20 were used for the determination of attitudes towards specific concepts involved in evaluation. This list is not exhaustive by any means, but represents some of the more important aspects to determine the benefits and drawbacks of evaluation as identified in the literature by Turner and Clift (1985). Each item was rated on a five point Likert scale, where 1 was strongly disagree, 2 was disagree, 3 - undecided, 4 represented agree and 5 indicated strongly agree.

A benefit is indicated by a mean rating of greater than three. Conversely, a drawback will be seen as an item which received a mean rating of less than three.

Informal Program Evaluation

The following section will present findings which relate to informal program evaluation. Mean values will be used to express group attitudes toward each of the 20 statements presented.

Informal evaluation received highest mean values for being a good idea (4.3), its ability to promote change in schools (4.3), and its ability to stimulate new ideas (4.2). Table 10 also shows the responses for informal evaluations ability to pinpoint strengths as well as weaknesses solidly in the agree range (4.2). A mean of 4.1 was observed for informal evaluation as a source of motivation for teachers and for coupling the evaluation of the program with the self-evaluation of the teacher (4.1). A good basis for questioning existing methods also received a mean of 4.1.

Table 10

Mean ratings for Informal Evaluation

Ungrouped	Mean	*n*
1. should include teachers observing other's lessons	3.9	76
2. can promote change in schools	4.3	79
3. should involve questionnaires	3.4	79
4. should assess parents' views	3.3	79
5. tends to make teachers feel threatened	2.9	79
6. should be coupled with self-evaluation	4.1	79
7. is, on the whole, a good idea	4.3	79
8. should be compulsory	3.1	79
9. should include a written report	3.0	79
10. is virtually a waste of time	1.8	79
11. can stimulate new ideas	4.2	78
12. should utilize student scores from provincial tests	2.4	79
13. should be done by administrators only	1.5	79
14. can be a source of motivation for teachers	4.1	79
15. provides a good basis for questioning existing methods	4.1	79
16. pinpoints strengths as well as weaknesses	4.2	79
17. should include taking pupils' views into account	3.8	79
18. is likely to be superficial	2.4	79
19. is too time consuming	2.6	79
20. should involve checklists	3.4	79

The observation of others lessons and taking into consideration pupils views received a high level of agreement with 3.9 and 3.8 respectively. A group of statements received mean values of just more than neutral in the range of 3.1 to 3.4. These statements dealt with aspects of methodology. The concepts of informal evaluation being a threat and including a written report both received undecided ratings.

Items which were rated with mild disagreement included informal evaluation being too time consuming (2.6) and superficial (2.4). Strong disagreement was recorded for evaluation being a waste of time, and for administrators only conducting informal evaluations.

Informal Program Evaluation Subgroup Differences

The data pertaining to informal evaluation have been again grouped into the three major subgroups; teaching experience, gender and school type. Significant differences are denoted in each table with an asterisk preceding the item number. Significance is determined at the .05 level.

Teaching Experience. Table 11 shows two significant differences among experience groups. The greatest deals with utilizing student scores from provincial tests in informal program evaluation. Teachers in their formative years responded with strong level of disagreement (1.7), while the most experienced group of teachers reported a mean of 2.6. This difference was significant at the .01 level. The likelihood of teachers agreeing to use student scores from provincial tests for informal program evaluation increases with years of experience.

Statement 1, dealing with teachers observing others lessons also received significantly different means ratings among the subgroups. Teachers with greatest experience were in slight agreement with a mean value of 3.5, while those in the midrange and least experience groups were strongly in agreement with means of 4.3 and 4.5 respectively.

None of the other statements produced differences which were significant.

Table 11

Means of Data: Informal Evaluation

Grouped by Teaching Experience (Years)	1 - 5	6 - 10	11+
n	16	20	59
*1. should include teachers observing other's lessons	4.5	4.3	3.5
2. can promote change in schools	4.2	4.3	4.3
3. should involve questionnaires	3.4	3.2	3.5
4. should assess parents' views	3.2	3.0	3.5
5. tends to make teachers feel threatened	3.0	2.9	2.9
6. should be coupled with self-evaluation	3.8	4.1	4.2
7. is, on the whole, a good idea	4.5	4.4	4.2
8. should be compulsory	3.5	3.0	3.0
9. should include a written report	2.9	2.9	3.0
10. is virtually a waste of time	1.7	1.5	1.9
11. can stimulate new ideas	4.5	4.3	4.2
*12. should utilize student scores from provincial tests	1.7	2.3	2.6
13. should be done by administrators only	1.4	1.5	1.6
14. can be a source of motivation for teachers	4.2	4.0	4.0
15. provides a good basis for questioning existing methods	4.1	4.0	4.1
16. pinpoints strengths as well as weaknesses	4.2	4.3	4.2
17. should include taking pupils' views into account	4.0	3.8	3.8
18. is likely to be superficial	2.2	2.3	2.5
19. is too time consuming	2.3	2.5	2.7
20. should involve checklists	3.3	3.7	3.3

Gender. Gender differences are easily visible in Table 12. Three of these differences were statistically significant. Again the statement involving student scores from provincial tests prompted large differences. Both gender groups responded in disagreement, but females responded with a lower mean score (2.2) than the males (2.8). As a group, females are more adverse to using provincial test

scores as an indicator in informal program evaluation. This difference is significant at the .01 level.

Table 12

Means of Data: Informal Evaluation

Grouped by Gender	F	M
n	66	28
1. should include teachers observing other's lessons	3.8	3.9
2. can promote change in schools	4.3	4.3
3. should involve questionnaires	3.4	3.5
4. should assess parents' views	3.3	3.3
5. tends to make teachers feel threatened	2.8	3.3
*6. should be coupled with self-evaluation	4.3	3.8
7. is, on the whole, a good idea	4.3	4.4
8. should be compulsory	3.0	3.3
9. should include a written report	2.9	3.2
10. is virtually a waste of time	1.8	1.9
11. can stimulate new ideas	4.2	4.3
*12. should utilize student scores from provincial tests	2.2	2.8
13. should be done by administrators only	1.6	1.5
14. can be a source of motivation for teachers	4.1	3.9
15. provides a good basis for questioning existing methods	4.1	4.0
16. pinpoints strengths as well as weaknesses	4.2	4.2
17. should include taking pupils' views into account	3.8	3.8
*18. is likely to be superficial	2.3	2.7
19. is too time consuming	2.6	2.7
20. should involve checklists	3.5	3.3

Strong differences in opinion between gender were reported for two further statements. Both groups responded that informal evaluation should be coupled with self-evaluation, but the female group to a substantially greater degree (4.3) than

their male counterparts (3.8). The difference for this statement was significant at the .03 level.

A similar response was given for informal evaluation being superficial, with females rejecting the idea to a greater degree than males. The mean value for male respondents was 2.7 compared to the female value of 2.3. This final finding of statistical significance for the gender subgroup ocurred at the .02 level.

School Type. Only one significant difference was determined to exist among the school subgroups. This difference was in response to the coupling of informal evaluation with self-evaluation. Table 13 depicts the large difference between junior and senior high schools responses to this concept. The junior high subgroup reaction was well into the strongly agree category with a value of 4.4, while senior high response was a firm agree with a mean of 3.7. This difference was significant at the .04 level.

Formal-Internal Program Evaluation

This section presents the findings of teacher attitudes as they relate to formal-internal program evaluation. Many of the mean scores for formal-internal evaluation demonstrate the same basic patterns as did informal evaluation, but with values which are not as extreme. In general statement responses have tended to move closer the neutral position.

Table 14 communicates the mean values for all respondents who stated having experience with formal-internal evaluation. As is evidenced by the change in *n* value from the previous evaluation type, fewer staff were involved in this form of evaluation.

Table 13

Means of Data: Informal Evaluation

Grouped by School Type	Elementary	Junior Senior High
n	56	15 26
1. should include teachers observing other's lessons	3.8	3.9 3.9
2. can promote change in schools	4.4	4.2 4.1
3. should involve questionnaires	3.5	3.3 3.3
4. should assess parents' views	3.3	3.5 3.2
5. tends to make teachers feel threatened	2.9	2.9 3.0
*6. should be coupled with self-evaluation	4.2	4.4 3.7
7. is, on the whole, a good idea	4.3	4.3 4.3
8. should be compulsory	3.2	2.8 3.0
9. should include a written report	3.0	2.8 3.0
10. is virtually a waste of time	1.8	2.0 1.7
11. can stimulate new ideas	4.3	4.5 4.0
12. should utilize student scores from provincial tests	2.2	2.4 2.8
13. should be done by administrators only	1.6	1.3 1.6
14. can be a source of motivation for teachers	4.1	4.2 3.8
15. provides a good basis for questioning existing methods	4.1	4.0 4.0
16. pinpoints strengths as well as weaknesses	4.1	4.2 4.3
17. should include taking pupils' views into account	3.8	3.9 3.8
18. is likely to be superficial	2.4	2.3 2.6
19. is too time consuming	2.5	2.4 3.0
20. should involve checklists	3.4	3.7 3.2

The highest mean ratings were given for the ability of formal-internal evaluation to pinpoint strengths as well as weaknesses, its ability to promote change, stimulate new ideas and being on the whole, a good idea. The lowest ratings were encourage by statements such as formal-internal evaluation being done by administrators only, being virtually a waste of time, and utilizing student scores

from provincial tests. The items which addressed superficiality and excessive time consumption also rated low mean scores.

Table 14

Mean Ratings for Formal-Internal Evaluation

Ungrouped	Mean	*n*
1. should include teachers observing other's lessons	3.7	58
2. can promote change in schools	4.1	57
3. should involve questionnaires	3.6	57
4. should assess parents' views	3.4	57
5. tends to make teachers feel threatened	3.2	57
6. should be coupled with self-evaluation	3.9	58
7. is, on the whole, a good idea	4.0	58
8. should be compulsory	3.3	58
9. should include a written report	3.8	58
10. is virtually a waste of time	2.2	58
11. can stimulate new ideas	4.0	58
12. should utilize student scores from provincial tests	2.4	58
13. should be done by administrators only	2.1	58
14. can be a source of motivation for teachers	3.8	58
15. provides a good basis for questioning existing methods	3.9	58
16. pinpoints strengths as well as weaknesses	4.2	58
17. should include taking pupils' views into account	3.7	58
18. is likely to be superficial	2.5	58
19. is too time consuming	2.8	58
20. should involve checklists	3.4	58

Formal-Internal Program Evaluation Subgroup Differences

Subgroup differences will be discussed only if they represent statistical significance. Mean values are used to express the overall opinion of the group.

Teaching Experience. Only one experience related difference was determined to be significant at the .05 level. This again dealt with item 12 and the use of student scores from provincial exams. Table 15 shows the distribution of mean score for the three experience groups.

Table 15

Means of Data: Formal - Internal Evaluation

Grouped by Teaching Experience (Years)	1 - 5	6 - 10	11 +
n	14	20	55
1. should include teachers observing other's lessons	3.8	3.8	3.7
2. can promote change in schools	4.1	3.8	4.2
3. should involve questionnaires	3.7	3.8	3.7
4. should assess parents' views	3.1	3.1	3.6
5. tends to make teachers feel threatened	3.2	3.8	3.0
6. should be coupled with self-evaluation	3.9	3.9	4.0
7. is, on the whole, a good idea	4.3	3.9	3.9
8. should be compulsory	3.9	3.1	3.2
9. should include a written report	4.1	3.9	3.8
10. is virtually a waste of time	2.1	2.2	2.3
11. can stimulate new ideas	4.0	3.8	4.1
*12. should utilize student scores from provincial tests	1.7	2.2	2.8
13. should be done by administrators only	2.6	2.1	2.0
14. can be a source of motivation for teachers	3.8	3.5	3.9
15. provides a good basis for questioning existing methods	4.1	3.5	4.0
16. pinpoints strengths as well as weaknesses	4.2	4.2	4.2
17. should include taking pupils' views into account	3.8	3.3	3.8
18. is likely to be superficial	2.6	2.7	2.4
19. is too time consuming	3.0	2.6	2.8
20. should involve checklists	3.8	3.8	3.3

The least experienced group are again most willing to reject this concept with a mean score of 1.7, while the teachers with most experience are less firm in their

disagreement at 2.8. Similarly to the case with informal evaluation, the teachers with medium experience responded between the two other groups.

Gender. Table 16 shows the distribution of means between the two gender subgroups. Only two of these differences are statistically significant.

Table 16

Means of Data: Formal - Internal Evaluation

Grouped by Gender	F	M
n	60	28
1. should include teachers observing other's lessons	3.8	3.5
2. can promote change in schools	4.2	4.0
3. should involve questionnaires	3.6	3.8
4. should assess parents' views	3.4	3.5
5. tends to make teachers feel threatened	3.1	3.5
6. should be coupled with self-evaluation	4.1	3.7
7. is, on the whole, a good idea	4.0	4.0
8. should be compulsory	3.4	3.1
9. should include a written report	3.9	3.7
*10. is virtually a waste of time	2.0	2.8
11. can stimulate new ideas	4.0	3.9
12. should utilize student scores from provincial tests	2.3	2.8
13. should be done by administrators only	2.1	2.2
14. can be a source of motivation for teachers	3.9	3.7
15. provides a good basis for questioning existing methods	4.0	3.7
16. pinpoints strengths as well as weaknesses	4.3	4.0
17. should include taking pupils' views into account	3.8	3.5
*18. is likely to be superficial	2.3	3.0
19. is too time consuming	2.6	3.2
20. should involve checklists	3.5	3.3

The difference with the higher level of significance was created by response to item 10. This states that formal-internal evaluation is virtually a waste of time. A .8

difference in mean was registered between females and males. Females were considerably stronger in their disagreement (2.0) than were the males at 2.8. This difference was significant at the .01 level.

Superficiality was the other concept which caused a gender mean difference which was determined to be significant at the .02 level. Male response was neutral (3.0), but females again reported a mean value consistent with stronger disagreement at 2.3.

In both of these cases the statements had negative connotations for formal-internal evaluation. A negative response to the negative wording in turn implies a positive response for program evaluation.

School Type. No significant difference was determined to exist in the mean values distributed among school types, as shown by Table 17.

Formal-External Program Evaluation

Formal-external evaluation uses set procedures and is controlled by a person external to the school. As a result, teachers attitudes toward this form of evaluation are again different than those of formal-internal evaluation. Table 18 demonstrates visually, that no mean values of four or higher exist. The other key factor, is that the number of participants who had experienced this form of evaluation diminished substantially to 34.

The highest means (3.9) were attributed to including a written report, and being coupled with self-evaluation. Respondents provided the second highest mean value of 3.8 for the ability of formal-external evaluation to stimulate new ideas and promote change in schools.

Table 17

Means of Data: Formal - Internal Evaluation

Grouped by School Type	Elementary	Junior	Senior High
n	50	13	26
1. should include teachers observing other's lessons	3.8	3.9	3.4
2. can promote change in schools	4.3	3.8	4.1
3. should involve questionnaires	3.8	3.6	3.6
4. should assess parents' views	3.5	2.9	3.6
5. tends to make teachers feel threatened	3.2	3.4	3.1
6. should be coupled with self-evaluation	4.0	4.1	3.7
7. is, on the whole, a good idea	3.9	4.4	3.9
8. should be compulsory	3.3	3.6	3.2
9. should include a written report	3.8	4.1	3.8
10. is virtually a waste of time	2.2	1.8	2.7
11. can stimulate new ideas	4.0	3.9	4.1
12. should utilize student scores from provincial tests	2.2	2.7	2.8
13. should be done by administrators only	2.1	2.3	2.0
14. can be a source of motivation for teachers	3.8	3.8	3.9
15. provides a good basis for questioning existing methods	3.9	3.9	3.9
16. pinpoints strengths as well as weaknesses	4.2	4.2	4.1
17. should include taking pupils' views into account	3.7	3.5	3.8
18. is likely to be superficial	2.5	2.2	2.7
19. is too time consuming	2.8	2.6	2.9
20. should involve checklists	3.5	3.7	3.2

Table 18

Mean Ratings for Formal-External Evaluation

Ungrouped	Mean	*n*
1. should include teachers observing other's lessons	2.9	34
2. can promote change in schools	3.8	34
3. should involve questionnaires	3.6	34
4. should assess parents' views	3.6	34
5. tends to make teachers feel threatened	3.4	34
6. should be coupled with self-evaluation	3.9	34
7. is, on the whole, a good idea	3.6	34
8. should be compulsory	2.9	34
9. should include a written report	3.9	33
10. is virtually a waste of time	2.7	34
11. can stimulate new ideas	3.8	34
12. should utilize student scores from provincial tests	2.3	34
13. should be done by administrators only	2.2	34
14. can be a source of motivation for teachers	3.6	34
15. provides a good basis for questioning existing methods	3.6	34
16. pinpoints strengths as well as weaknesses	3.7	34
17. should include taking pupils' views into account	3.6	34
18. is likely to be superficial	2.8	34
19. is too time consuming	3.0	34
20. should involve checklists	3.3	34

The statements which received the lowest means of 2.2 and 2.3 respectively involved the evaluation being done by administrators only and utilizing student scores for provincial tests in formal-external evaluation.

Formal External Program Evaluation Subgroup Differences

Similar issues were again responsible for the significant subgroup differences which are reported in this section. The number of participants within each subgroup has again decreased, but in proportion to the overall decrease in participant experience with formal-external evaluation.

Teacher Experience. One single difference was significant among the experience groups. This again relates to the use of students scores from provincial tests as an indicator for program evaluation. Table 19 demonstrates the same trend which has been seen for the other subclassifications of evaluation as well. The striking difference in means between the least experience group and the most experience group is significant at the .01 level. Both values indicate a disagreement with the statement, but the mean of 1.3 for the teachers with five or less years a much stronger degree of opposition to the concept than the 2.6 value of the veteran teachers.

Gender. As Table 20 demonstrates, one significant difference in gender was found to exist and again dealt with the issue of program evaluation likely being superficial. Gender responses are on opposing sides of the issue with males reporting strong agreement (3.9), and females providing a mean of 2.4. This difference is significant at the .01 level.

Table 19

Means of Data: Formal - External Evaluation

Grouped by Teaching Experience (Years)	1 - 5	6 - 10	11 +
n	4	7	23
1. should include teachers observing other's lessons	3.8	3.1	2.6
2. can promote change in schools	4.0	4.1	3.6
3. should involve questionnaires	3.3	3.9	3.6
4. should assess parents' views	3.5	3.6	3.6
5. tends to make teachers feel threatened	3.3	3.6	3.4
6. should be coupled with self-evaluation	4.3	3.9	3.9
7. is, on the whole, a good idea	4.3	3.7	3.4
8. should be compulsory	3.3	2.3	3.0
9. should include a written report	4.3	3.7	3.9
10. is virtually a waste of time	2.3	2.7	2.7
11. can stimulate new ideas	4.3	4.1	3.6
*12. should utilize student scores from provincial tests	1.3	2.0	2.6
13. should be done by administrators only	2.3	1.6	2.3
14. can be a source of motivation for teachers	4.3	3.7	3.4
15. provides a good basis for questioning existing methods	3.8	3.9	3.5
16. pinpoints strengths as well as weaknesses	3.8	3.7	3.7
17. should include taking pupils' views into account	3.8	3.7	3.6
18. is likely to be superficial	2.5	3.0	2.8
19. is too time consuming	3.0	3.3	2.9
20. should involve checklists	3.5	3.6	3.1

Table 20

Means of Data: Formal - External Evaluation

Grouped by Gender	F	M
n	24	10
1. should include teachers observing other's lessons	3.0	2.4
2. can promote change in schools	3.9	3.4
3. should involve questionnaires	3.7	3.3
4. should assess parents' views	3.6	3.4
5. tends to make teachers feel threatened	3.3	3.9
6. should be coupled with self-evaluation	4.1	3.5
7. is, on the whole, a good idea	3.8	3.1
8. should be compulsory	3.0	2.6
9. should include a written report	4.0	3.6
10. is virtually a waste of time	2.3	3.5
11. can stimulate new ideas	4.0	3.2
12. should utilize student scores from provincial tests	2.3	2.2
13. should be done by administrators only	2.1	2.4
14. can be a source of motivation for teachers	3.7	3.3
15. provides a good basis for questioning existing methods	3.9	2.9
16. pinpoints strengths as well as weaknesses	3.9	3.2
17. should include taking pupils' views into account	3.6	3.7
*18. is likely to be superficial	2.4	3.9
19. is too time consuming	2.7	3.6
20. should involve checklists	3.4	3.0

School Type. Table 21 shows that although many differences exist among the attitudes which were reported by the subgroups, only three were determined to be significant. The concept of a written report provided a mean value of 4.2 from elementary teachers, but only 3.5 from the senior high group. This difference was significant at the .04 level.

Table 21

Means of Data: Formal - External Evaluation

Grouped by School Type	Elementary	Junior Senior High
n	18	3	13
1. should include teachers observing other's lessons	2.9	1.7	3.1
2. can promote change in schools	4.0	2.7	3.7
3. should involve questionnaires	3.7	3.0	3.6
4. should assess parents' views	3.6	3.0	3.6
5. tends to make teachers feel threatened	3.3	4.0	3.5
6. should be coupled with self-evaluation	4.1	4.0	3.6
7. is, on the whole, a good idea	3.8	3.3	3.4
8. should be compulsory	2.8	2.7	2.9
*9. should include a written report	4.2	4.0	3.5
10. is virtually a waste of time	2.4	3.3	2.8
11. can stimulate new ideas	4.0	3.3	3.5
*12. should utilize student scores from provincial tests	2.0	2.0	2.8
*13. should be done by administrators only	2.2	3.3	1.9
14. can be a source of motivation for teachers	3.9	3.3	3.2
15. provides a good basis for questioning existing methods	3.7	3.3	3.5
16. pinpoints strengths as well as weaknesses	3.9	3.7	3.3
17. should include taking pupils' views into account	3.5	3.0	4.0
18. is likely to be superficial	2.7	3.3	2.8
19. is too time consuming	2.8	3.0	3.2
20. should involve checklists	3.5	3.3	2.9

Administrators only evaluating programs also provided a wide difference in means, but involving junior and senior high groups this time. Junior high responded slightly on the agree side of neutral (3.3), while the senior high group mean was 1.9, representing disagreement with the concept. This difference was significant at the .04 level.

Again the concept of using student scores from provincial tests created a significant difference in response. Elementary and junior high teachers responded with means of 2.0 while the senior high mean was closer to neutral at 2.8. All responses indicate disagreement with the statement, but the degree is stronger in the elementary and junior high groups. This finding is significant at the .05 level.

Comparison of Means by Evaluation Subclassification

Significance of the difference in attitude responses for evaluation subclassification was not tested. The comparison is made only to assist in the understanding of general trends which appear within the data. Table 22 compares the overall mean scores of the sample across the subclassifications of evaluation.

Item numbers 3, 4, 6, 8, 12, 17,18, 19, 20 show little difference in means among the types of evaluation.

Statement 1 shows a .9 point difference between the two categories of internal evaluation, when compared with formal-external evaluation. This shows that the observation of lessons is accepted for either formal or informal internal evaluation while considered slightly negative for evaluations that have external control.

The response to item 2 (4.3) shows strongly that it is considered that informal evaluation can promote change in schools. The mean for formal-external evaluation scored .5 lower showing a slightly less positive attitude toward the potential for promotion of change. These results point to a generally positive attitude toward program evaluation for promoting change in schools.

Item 5 indicates that in general, teachers feel very little threat from program evaluation. Informal evaluation received a neutral mean of 2.9 while the formal

external received a slightly positive score of 3.4. That indicates only a slight threat overall when evaluation control is external.

Table 22

Comparison of Means by Subclassification of Evaluation Type

Evaluation type.	IE	FI	FE
n	79	58	34
1. should include teachers observing other's lessons	3.8	3.7	2.9
2. can promote change in schools	4.3	4.1	3.8
3. should involve questionnaires	3.4	3.7	3.6
4. should assess parents' views	3.3	3.4	3.6
5. tends to make teachers feel threatened	2.9	3.2	3.4
6. should be coupled with self-evaluation	4.1	3.9	3.9
7. is, on the whole, a good idea	4.3	4.0	3.6
8. should be compulsory	3.1	3.3	2.9
9. should include a written report	3.0	3.8	3.9
10. is virtually a waste of time	1.8	2.2	2.7
11. can stimulate new ideas	4.2	4.0	3.8
12. should utilize student scores from provincial tests	2.4	2.4	2.3
13. should be done by administrators only	1.5	2.1	2.2
14. can be a source of motivation for teachers	4.1	3.8	3.6
15. provides a good basis for questioning existing methods	4.1	3.9	3.6
16. pinpoints strengths as well as weaknesses	4.2	4.2	3.7
17. should include taking pupils' views into account	3.8	3.7	3.6
18. is likely to be superficial	2.4	2.5	2.8
19. is too time consuming	2.6	2.8	3.0
20. should involve checklists	3.4	3.4	3.3

Note. IE = Informal evaluation; FI = Formal-internal; FE = Formal-external.

The mean values associated with statement 7, evaluation being on the whole a good idea, points to a more positive attitude (4.3) toward informal evaluation, than to external evaluation (3.6). Formal-internal also received a positive response at 4.0.

The creation of a written report received means indicating positive support, but only in a formalized evaluation. Responses for informal evaluation were neutral.

Informal evaluation was not considered a waste of time (1.8). As the type of evaluation became more formalized or controlled, the degree of similarity to this statement lessened. Formal-external evaluation received a 2.7, just slightly removed from the neutral position of 3.0.

There was positive support for the notion of evaluation stimulating new ideas. The degree to which the means suggest agreement varied by evaluation type, but the more control was teacher centered, the more the means indicated agreement for stimulating new ideas.

The concept of administrators only, doing the evaluations of programs received a negative response. The negative response was stronger with informal evaluation than with formal-external evaluation. That indicates the involvement of administrators is more acceptable in formalized evaluation, but not to the exclusion of other persons.

Teachers responded more positively to informal evaluation (4.1) being a source of motivation than either formal-internal (3.8) or formal-external (3.6) evaluations. Again the trend was more positive for formal-internal than for formal-external evaluation.

As far as evaluation providing a good basis for questioning existing methods, teachers responded positively to informal evaluation (4.1), slightly less positive with formal-internal (3.8) and just slightly less again for formal external evaluation (3.6).

The ability of evaluation to pinpoint strengths and weaknesses was supported by teachers mean responses of 4.2 for informal evaluation, 4.2 for formal-internal evaluation and 3.7 for formal-external evaluation.

General Impressions of Program Evaluation

Subproblem three involves the determination of general attitudes toward program evaluation in relation to selected issues of implementation. The section section of the instrument used for collecting these attitudes did not make a distinction regarding the evaluation experience of respondents.

It was deemed important to determine just how the general sample felt about several important issues which contribute the the effectiveness of program evaluation used as a tool within social institutions. Table 23 shows the issues and the mean responses which were collected. The strongest, most cohesive response was that to be involved in program evaluation, teachers need to feel their effectiveness will be improved in some manner. In addition, the respondents felt that teachers would make competent evaluation coordinators. A strong response also indicated that teachers prefer to have the control of program evaluation within the school as opposed to external. It was also generally held that program evaluation does provide useful information for staff, upon which to base decisions.

There was moderate agreement that teachers needed more information about program evaluation. Participants responded with slight negativity toward Alberta Education making program evaluation mandatory. Respondents are unsure if the

recommendations of program evaluations are implemented. All of these points tend to demonstrate the slight threat that is felt toward program evaluation as a concept.

Table 23

Means - General Impressions of Program Evaluation

Ungrouped	Mean	*n*
1. Recommended changes are rarely implemented	3.0	95
2. Teachers can be competent evaluation coordinators	4.2	95
3. Program evaluation provides staff with useful information upon which to make decisions	4.0	94
4. Alberta Education should mandate program evaluation	2.7	94
5. Teachers need to feel that their effectiveness will be improved in some way as a result of being involved in program evaluation	4.4	95
6. More information, debate and seminars on program evaluation are required	3.5	95
7. School-based control of program evaluation is preferable to outside control	4.0	95

General Impressions Subgroup Differences

Using the same subgroups as before, an attempt to determine significant differences was undertaken. The distribution of attitudes among the various subgroups are displayed in the tables to follow.

Teaching Experience. An analysis of variance for experience subgroups found no significant differences. Table 24 shows the degree of similarity of means.

Table 24

Means - General Impressions of Program Evaluation

Grouped by Teaching Experience (Years)	1 - 5	6 - 10	11+
n	13	16	50
1. Recommended changes are rarely implemented	2.9	2.9	3.0
2. Teachers can be competent evaluation coordinators	4.2	4.2	4.2
3. Program evaluation provides staff with useful information upon which to make decisions	4.0	3.9	3.9
4. Alberta Education should mandate program evaluation	2.8	2.9	2.6
5. Teachers need to feel that their effectiveness will be improved in some way as a result of being involved in program evaluation	4.5	4.4	4.4
6. More information, debate and seminars on program evaluation are required	3.5	3.6	3.5
7. School-based control of program evaluation is preferable to outside control	4.1	4.1	4.0

Gender. Table 25 reveals the only significant gender difference to center around the issue of recommendations not being implemented. Males are on the positive side while females are on the negative. This implies that males are not as sure that recommendations would be implemented while females are are more confident they would be realized. This difference is significant at the .01 level.

School Type. Table 26 shows the distribution of mean values by category of school type. No significant differences exist.

Table 25

Means - General Impressions of Program Evaluation

Grouped by Gender	F	M
n	56	23
*1. Recommended changes are rarely implemented	2.8	3.4
2. Teachers can be competent evaluation coordinators	4.2	4.1
3. Program evaluation provides staff with useful information upon which to make decisions	4.1	3.7
4. Alberta Education should mandate program evaluation	2.7	2.7
5. Teachers need to feel that their effectiveness will be improved in some way as a result of being involved in program evaluation	4.4	4.4
6. More information, debate and seminars on program evaluation are required	3.6	3.3
7. School-based control of program evaluation is preferable to outside control	4.0	4.0

Attitude Categories

Subproblem four deals with the factor analysis of the collected data, comparing the results with those found in the Solihull research of Turner and Clift (1985). The Solihull study dealt with formal-internal evaluation only, and therefore comparisons here will only deal with the data acquired from the sections involving informal and formal-internal evaluation. The results pertaining to informal evaluation may not be as applicable for comparison but is included for contrast.

The procedure used by Turner and Clift (1985) to analyze data was followed in order that reasonable comparisons could be made. One distinction was that interpretation of the data in this study was done with factor loadings of .4 and

greater, rather than the .3 and greater used by Turner and Clift. This was to eliminate some of the cross-loading with other factors which takes away from the clarity of the findings. Seeking only major factors rather than discovering all factors allows for this difference.

Table 26
Means - General Impressions of Program Evaluation

Grouped by School Type	Elementary	Junior	Senior
n	50	10	19
1. Recommended changes are rarely implemented	2.8	3.3	3.2
2. Teachers can be competent evaluation coordinators	4.2	3.8	4.3
3. Program evaluation provides staff with useful information upon which to make decisions	4.0	3.9	3.7
4. Alberta Education should mandate program evaluation	2.5	2.8	3.0
5. Teachers need to feel that their effectiveness will be improved in some way as a result of being involved in program evaluation	4.5	4.3	4.3
6. More information, debate and seminars on program evaluation are required	3.6	3.2	3.4
7. School-based control of program evaluation is preferable to outside control	4.1	3.8	3.9

Formal-Internal Evaluation Factors

Comparison of these findings with the Solihull study are logical. Both studies involved attitudes attained based on the same classification of evaluation. A full 62.8% of the variance in attitudes was described by these five factors. Turner and Clift were able to attribute 58% of the variance with their five factors.

The concept of professionality was clearly determined to exist as a separate factor with formal-internal evaluation as shown in Table 27. It also appeared in the Solihull study, but explained much less of the variance in attitudes.

Table 27

Factor One: Professionality / Professional Development (19.9%)

#	Statement	Loading
	Program evaluation:	
C14	can be a source of motivation for teachers	.85
C15	provides a good basis for questioning existing methods	.84
C11	can stimulate new ideas	.80
C2	can promote change in schools	.64
C7	is on the whole a good idea	.58
C16	pinpoints strengths as well as weaknesses	.53

Efficiency was included with threat in the Solihull study, but both appeared as distinct factors here. In this study they appeared side by side as factors two and three. Table 28 shows the relationship of the concept variables for efficiency and Table 29 provides the statements for determination of the threat factor.

Collegiality and openness appear in both studies to act as a single factor. It could indicate that the two issues are intertwined. Table 30 demonstrates the grouping of statements which represent these concepts.

Table 28

Factor Two: Efficiency (14.2%)

#	Statement	Loading
	Program evaluation:	
C18	is likely to be superficial	-.74
C19	is too time consuming	-.74
C10	is virtually a waste of time	-.71
E6	More information, debate and seminars on program evaluation are required	-.68

Table 29

Factor Three: Threat (7.57%)

#	Statement	Loading
	Program evaluation:	
C12	should utilize student scores from provincial tests	.73
C6	should be coupled with self-evaluation	-.69
C5	tends to make teachers feel threatened	.54

Table 30

Factor Four: Collegiality / Openness (10.7%)

#	Statement	Loading
	Program evaluation:	
C4	should assess parents' views	.66
C17	should include taking pupils' views into account	.61
C1	should include teachers observing other's lessons	.57
C13	should be done by administrators only	-.55

Formality was again determined to be a single factor in both studies. The appearance of this tends to follow logically, as both studies used attitudes based on a formal evaluation. Table 31 shows the concept statements.

Table 31

Factor Five: Formality (9.6%)

#	Statement	Loading
	Program evaluation:	
C8	should be compulsory	.82
C9	should include a written report	.50
E4	Alberta Education should mandate program evaluation	-.59

In summary, five factors have been determined to exist in the attitudes collected in the sample. These attitudes were derived based upon actual experience with formal-internal program evaluation. The factors involved deal with the same concepts which Turner and Clift (1985) determined to exist, with one exception. This is the factor of "professional development" which appears to be interpreted as "career advancement" in Britain. During the creation of the instrument, the term "professional development" was interpreted in the Canadian sense, and variables assigned accordingly. This interpretation eliminated the possibility that this factor might be detected in the British context.

The other major difference was the separation of the issues of threat and efficiency into distinct factors. This verifies their existence in the sample, but does not connect them as a single entity.

Informal Program Evaluation Factors

It is important to readdress the notion that although the name does not explicitly state, informal program evaluation is a form of internal evaluation. This allows for the logical comparison of informal program evaluation findings with those of formal-internal evaluation.

The following major factors were found to account for 52.2% of the variance in attitudes toward informal program evaluation. Included in the factor determination are only the responses from those people who indicated some participation in this type of evaluation.

Informal evaluation was determined to have seven underlying factors, but only the first six included sufficient variables to determine appropriate labels. The factors for this classification of evaluation were not precise, indicating the presence of underlying minor factors as well. It is important to note, that although the results of this study for informal evaluation have been compared with the results of the Solihull study, it is for the purposes of drawing conclusions only. Logic would advise that similarity or differences of attitude factors between informal evaluation and formal-internal evaluation may be attributable to many factors. A shift in the formality of the process is the large variable.

Factor one for informal evaluation appears to be a new or different factor not discovered in the Solihull study or in this present study of formal-internal evaluation. Four of the seven variables in this category represent either collegiality or openness and are denoted with asterisks in Table 32. The remainder tend to represent efficiency. Because openness and efficiency do not conceptually fit, these must represent a different factor involved in informal evaluation which is not present in formal evaluation.

Table 32

Factor One: Efficiency / Openness (13.7%)

#	Statement	Loading
	Program evaluation:	
B10 *	is virtually a waste of time	-.81
B11	can stimulate new ideas	.71
B19	is too time consuming	-.60
B17 *	should include taking pupils' views into account	.55
B5	tends to make teachers feel threatened	-.55
B13 *	should be done by administrators only	-.53
B7 *	is on the whole a good idea	.47

With formal-internal evaluation, threat also factored out as s separate entity, unlike the combined efficiency and threat in Turner and Clift's results. In informal evaluation, the concept of efficiency was found to load in the first category with openness. Table 33 shows the statements and loadings for the threat factor.

Table 33

Factor Two: Threat (7.1%)

#	Statement	Loading
	Program evaluation:	
B12	should utilize student scores from provincial tests	-.80
B6	should be coupled with self-evaluation	-.73

The concept of formality was found to exist as a separate factor again, just as in the formal-internal evaluation results as well as Turner and Clift's study. Table 34 shows the statements which determined this factor.

Table 34

Factor Three: Formality (8.2%)

#	Statement	Loading
	Program evaluation:	
B4	should assess parents' views	.81
B9	should include a written report	.75

Professionality contributed strongly to the explanation of the total variance which existed. Informal evaluation is quite certainly an activity which is seen as being professional conduct. This finding is similar to the case of formal-internal evaluation. Table 35 demonstrates the relationship of the variables involved.

Table 35

Factor Four: Professionality (12.4%)

#	Statement	Loading
	Program evaluation:	
B15	provides a good basis for questioning existing methods	.81
B14	can be a source of motivation for teachers	.63
B18	is likely to be superficial	-.63
B16	pinpoints strengths as well as weaknesses	.57
B8	should be compulsory	.51
B19	is too time consuming	-.42

The term professional development was determined to have a drastically different usage in Canada than in Britain. Canadians see professional development as growth as a professional, while the British tend to use it for the description of career advancement. Although these issues are not comparable, the statements

provided in Table 36 indicate a separation from the factor of professionality /professional development as it occured in formal-internal evaluation.

Table 36

Factor Five: Professional Development / Growth (8.8%)

#	Statement	Loading
E5	Teachers need to feel that their effectiveness will be improved in some way as a result of being involved in program evaluation	-.84
	Program Evaluation:	
B2	can stimulate new ideas	.64
B6	should be coupled with self-evaluation	.47
B13	is virtually a waste of time	-.43

Informal evaluation attitudes of this sample created a separate factor of collegiality. This was neither the case in the Solihull study, where it was grouped with openness, nor in the formal-internal assessment discussed previously. Table 37 shows the variables involved in the creation of this factor.

In general the results generated by informal evaluation attitudes and those created from formal-internal program evaluation cannot be favorably compared. The lack of clarity generated in the factors and the differences in categories determined suggest incompatibility. The cause of this incompatibility is most likely the difference in subclassification of evaluation type. That is the lack of "formal" evaluation attitudes.

Table 37

Factor Six: Collegiality (9.2%)

#	Statement	Loading
	Program evaluation:	
B2	should include teachers observing other's lessons	.78
B9	should be made compulsory for schools	.48
E4	Alberta Education should mandate program evaluation	-.78

The preceding discussions have shown only the variables with loadings of .4 or greater. The complete matrix of rotated factor loadings is presented in Appendix 3.

Chapter Five

Summaries, Conclusions, Implications and Recommendations

The primary outcomes of this research are the description of a) the extent to which staff within the Calgary Board of Education have been involved with program evaluation, b) the perceived benefits and drawbacks for being involved with the different classifications of program evaluation and c) the major underlying factors involved with internal program evaluation and the comparison of these factors with those derived by Turner and Clift (1985).

Summaries and Conclusions

Summaries of the findings and related conclusions for each subproblem are presented in this section. Subproblems one and two also address the elements of control and formality within the three subclassifications informal, formal-internal and formal-external program evaluation.

Subproblem One - Summary

Subproblem one read as follows:

> **To what extent have staff been involved in "evaluating" school programs, either formally or informally?**

The extent of staff involvement with program evaluation is varied. Only 6.2% of the sample reported having no involvement in program evaluation of any kind. Generally, involvement in program evaluation decreases as the structure of the evaluation is increased, and decreases further with the addition of external control. This latter decrease may be in part attributable to the expense of external evaluations, and therefore the resultant infrequency with which they occur.

Informal program evaluation. The frequency of non-participation in informal evaluation was 16.8% of the sample. This indicates 83.2% of staff have been involved to some extent in informal program evaluation. The analysis of involvement in informal program evaluation by years of teaching experience, seems to reveal a trend towards more activity in recent years. If one program evaluation per year of teaching experience is used as a guideline for comparison of this data, the highest percentage involvement is in the group with the least experience, followed by those with the most experience, and the moderate experience group last. Table 38 shows this trend with the comparison information underlined. This may be an indicator that emphasis and the perceived importance of informal program evaluation has changed over time. It would seem that there may have been a recent increase in informal program evaluation activity.

Table 38

Percentage Involvement in Informal Program Evaluation by Teaching Experience

	Years of Experience			
Program Evaluations	1-5	6-10	11+	TOTAL
n	14	20	55	
0	18.8	20.0	15.3	16.8
1-2	<u>31.2</u>	25.0	32.2	30.5
3-5	<u>25.0</u>	40.0	20.3	25.3
6-10	12.5	<u>5.0</u>	13.6	11.6
11+	12.5	10.0	<u>18.6</u>	15.8
Total %	100	100	100	100

<u>Note</u>. Underlined data suggests one evaluation per year of experience.

Formal-internal program evaluation. The percentage of involvement in formal-internal evaluation was 64%. This is a reduction from that of informal evaluation, but still a fairly high percentage. Involvement by teaching experience exhibits a trend similar to that which occurred for informal evaluation. Teachers with the least teaching experience have had the greatest percentage involvement in formal-internal program evaluation. Table 39 illustrates this trend.

Table 39

Percentage Involvement in Formal-Internal Program Evaluation by Teaching Experience

	Years of Experience			
Program Evaluations	1-5	6-10	11+	TOTAL
n	14	20	55	
0	35.7	35.0	36.4	36.0
1-2	50.0	45.0	45.5	46.1
3-5	14.3	10.0	10.9	11.2
6-10	0.0	5.0	5.4	4.5
11+	0.0	5.0	1.8	2.2
Total %	100	100	100	100

Note. Underlined data suggests one evaluation per year of experience.

Formal-external program evaluation. The same trend is not true for formal-external evaluation. Involvement in this type of evaluation appears to vary somewhat proportionally with experience. Those teachers with the most teaching experience having had the greater experience with formal-external evaluation, while

those with the least teaching experience have fewer experiences of this type. Additionally, those teachers in senior high school have considerably more experience with formal-external evaluation than other teachers. Overall, 42% of teachers have had some involvement with externally controlled program evaluation.

Subproblem One - Conclusions

It may be concluded that experience with program evaluation for this sample is diverse. Involvement in informal program evaluation is greater than the other subclassifications. This is logical realizing that informal evaluation requires no formalized process, and tends to be completed on an individual basis requiring a minimum amount of time. The results are generally used only by the individual involved, as an effort to fine-tune a program.

The analysis of experience with informal program evaluation by years of teaching provides the insight that the least experienced group had a higher percentage involvement than the other groups. This may lead to the logical conclusion that within the last five years, it has become more expected or accepted, to evaluate and revise the programs which are taught.

Subproblem Two - Summary

Subproblem two read as follows:

What are the perceived benefits and drawbacks of program evaluation?

Benefits and drawbacks have been determined strictly from the extent to which respondents agreed with specific attitude statements derived from Turner and

Clift (1985). These results then are not exhaustive, but conclusions may be drawn based upon the information provided. The benefits and drawbacks will be summarized here by subclassification of evaluation type. The significance level for this study was set at .05. Factors which caused a variability greater than this were taken to be statistically significant.

Informal program evaluation. The mean scores for informal program evaluation provide us with varying degrees of positive response to all but two of the variable statements. Teachers rated the creation of a written report as undecided. The thought of using student scores from provincial tests received 2.4, the only mean less than the midpoint. The most positive aspects of informal program evaluation deal with the ability to promote change within the school, pinpoint strengths and weaknesses and stimulate new ideas. As well, among these most positive included the ability to be a source of motivation for teachers, providing a good basis for questioning existing methods, being coupled with self-evaluation and being on the whole, a good idea. These may all be seen as clear benefits for participation in informal program evaluation.

Significant group differences were found to exist within all three subgroups. By teaching experience, there were two significant issues. Disagreement with the use of provincial test scores was significantly less in the group of teachers with most experience when compared with the least experienced group. This difference was significant at the .01 level. A reverse trend was seen with the issue of teachers observing others lessons. Here the group with least experience was most positive, while the most experienced group was closer to neutral. This difference was significant at the .03 level.

In the gender subgroup, there were three significant differences. Most significant at the .01 level was the male / female difference about using student scores from provincial tests. Both groups rated the statement in the disagree range, but the female group was more strongly opposed to this statement. A similar response was received for the idea of informal evaluation being superficial. Both groups rated it with disagreement, but females to a greater degree than males. This finding was significant at the .02 level. The last difference is the one created by the assertion that informal program evaluation should be coupled with self-evaluation. The mean female response was of strong agreement, while the male group was only slightly in agreement. This difference was significant at the .03 level.

The school type subgroup revealed only one significant difference. This difference was brought on by the statement which suggests that informal program evaluation should be coupled with self-evaluation. Responses by the junior and senior high groups represented the greatest difference. Junior high teachers responded with a mean value representing strongly agree, while the senior high value came close to the agree mark. Elementary response was in the middle. This difference was significant at the .04 level.

Formal-internal program evaluation. The mean scores show to a lesser degree, the same basic positive points which were expressed in the discussion on informal program evaluation. The noteworthy features which are specific to formal-internal evaluation will be summarized here. The statements which received the strongest affirmation among the positive points involved the ability to pinpoint strengths as well as weaknesses, to promote change in schools and stimulate new ideas. These were also the three statements rated most highly in conjunction with informal program evaluation.

With formal-internal evaluation two concepts were rated as being negative. They involve teachers perceptions that this form of program evaluation is slightly threatening, and again reveals the dislike which is expressed for using student scores from provincial tests as an indicator of program quality.

Three significant subgroup differences were ascertained to exist. Two of these differences involved grouping according to the amount of teaching experience. The major issue again dealt with the use of students scores from provincial tests. The group with most experience expressed only slight disagreement, while their least experienced colleagues responded with a mean score representing strong disagreement. This difference was significant at the .05 level.

The other two differences are gender group related. Both groups demonstrated that formal-internal evaluation was not considered to be a waste of time, but females to a greater degree than males. A .8 difference on a five point scale was found to be significant at the .01 level. Along the same line, the concept of formal-internal evaluation being superficial created a male response of neutral, while females asserted firm disagreement. This difference was significant at the .02 level.

Formal-external program evaluation. This subclassification of evaluation received no mean values between four and five, indicating again a slight lowering in the overall positive attitudes toward this approach. Again, ranking among the highest mean values were the concepts of promotion of change in schools and the stimulation of new ideas. The strongest mean value was provided for the creation of a written report.

The weakest mean values were attributed to the statements involving evaluation being conducted by administrators only and the use of student scores on provincial tests. These two statements could be categorized as unpopular at best.

Five significant subgroup differences were determined to exist for formal-external evaluation. Teaching experience again provided a strong difference between the polar groups for the concept of using student scores from provincial tests. The junior group provided a mean indicative of strong disagreement, as opposed to the senior group who found the idea much less objectionable. This difference was significant at the .01 level.

Significant gender differences were limited to the image of formal-external evaluation being superficial. Males provided a mean rating indicating agreement with the idea while females responded with moderate disagreement. This difference was significant at the .04 level.

The three other significant differences occurred among the school type subgroups. The concept of a written report created a difference between the elementary and senior high groups which was significant at the .04 level. Elementary teachers were far more in favor of creating the report than were senior high teachers. A difference which centered around administrators only evaluating programs, saw the junior high group with a mean value registering slight agreement, while the senior high mean representing disagreement. This difference was significant at the .04 level. Lastly, the concept of using student scores from provincial tests created a significant split between the senior high mean and the elementary and junior high means. In all cases the value represented disagreement with the concept, but the elementary and junior high groups it was a stronger

disagreement than with the senior high group. This difference was significant at the .05 level.

One possible reason for the differences which were found to occur between the senior high and the other school groups regarding the use of student scores from provincial tests may be attributed to the fact that high school is extremely content oriented. Other levels of school spend a great deal of teaching time on social issues and classroom management. High school teachers are able to spend more of their time assuring that students are ready for the test.

To clarify and summarize the issues which received mean values in excess of .4 on either side of the neutral value of 3, Table 40 was created. It provides a simplified view of what might be called the benefits and drawbacks of program evaluation. This table does not show the magnitude of the mean value, just the indicated direction toward benefit or drawback. Table 40 allows for the comparison of benefits and drawbacks between subclassifications of programs evaluation.

Informal evaluation and formal-external evaluation were determined to have eleven statements rated as benefits while formal-internal evaluation received thirteen such statements. Eight of the statements were rated as benefits in all three subclassifications of program evaluation. All three subclassifications of program evaluation were determined to have the same two statements rated as drawbacks; that program evaluation be done by administrators only and that program evaluation should utilize student scores from provincial tests.

Table 40

Benefits and Drawbacks Compared by Evaluation Subclassification

	Informal		Formal-Internal		Formal-External	
	B	D	B	D	B	D
1. should include teachers observing other's lessons	√		√			
2. can promote change in schools	√		√		√	
3. should involve questionnaires			√		√	
4. should assess parents' views					√	
5. tends to make teachers feel threatened						
6. should be coupled with self-evaluation	√		√		√	
7. is, on the whole, a good idea	√		√		√	
8. should be compulsory						
9. should include a written report			√		√	
10. is virtually a waste of time	√	(X)	√	(X)		
11. can stimulate new ideas	√		√		√	
12. should utilize student scores from provincial tests		X		X		X
13. should be done by administrators only		X		X		X
14. can be a source of motivation for teachers	√		√		√	
15. provides a good basis for questioning existing methods	√		√		√	
16. pinpoints strengths as well as weaknesses	√		√		√	
17. should include taking pupils' views into account	√		√		√	
18. is likely to be superficial	√	(X)	√	(X)		
19. is too time consuming						
20. should involve checklists						

Note. B = Benefit; D = Drawback; √= mean value >3.4; X = mean value < 2.6. Statements 10 and 18 display "X" values in parentheses, which indicate disagreement with a negative statement. This represents a "positive", or lack of drawback for the subclassification of program evaluation indicated.

Subproblem Two - Conclusions

In general, there are more benefits indicated than drawbacks for all types of program evaluation. The number of benefits and drawbacks are not as important as the concepts which these items describe. Although the conceptual framework for this study will be discussed later in this chapter, all of the benefit statements tend to point to a general sense of professionality and faith in the ability of the processes to assist in program revision, school change and growth. Distinct problems would occur if administrators only, were involved in the evaluation process, or if the student scores from provincial tests were used as an indicator of effective programs.

Assessing parents views were rated to be beneficial only in a formal-external evaluation. This implies that parents views are not important for internally controlled evaluations. A logical conclusion may be drawn that parents views represent accountability, and that staff associate the use of formal-external evaluation with accountability to a greater extent than the other subclassifications of program evaluation. Another way of looking at this occurrence, is to conclude that parents views are not as likely to be important for the purposes of program improvement.

Several issues maintained mean values close to the neutral or undecided position. These included the time element needed to accomplish the task of program evaluation, the use of checklists, if the processes makes teachers feel threatened, and if the process should become compulsory.

Subproblem Three - Summary and Conclusions

Subproblem three reads as follows:

Determine the views of all respondents on selected issues relating to the benefits and drawbacks of implementing program evaluation.

The issues to which participants responded in this section dealt primarily with "general impressions" of program evaluation. These selected issues were rated by all participants, not just those with previous evaluation experience. The intent was to determine what type of issues might be at the heart of encouraging or discouraging widespread participation in program evaluation. Four of the seven statements were derived from Turner and Clift (1985). The other three were determined from discussions held with program consultants.

Two of theses issues received near neutral mean values while the other five issues received mean values which suggested agreement by exceeding a mean rating of 3.4 on a five point scale. The strongest of these issues dealt with the need of teachers to feel that their effectiveness would be improved in some way as a result of being involved in program evaluation. It is evident that if teachers were willingly going to participate in program evaluations, they must see that the process would benefit them professionally. Findings from subproblem two tend to indicate that teachers already see program evaluation as providing just these benefits.

The second strongest issue was that teachers can be competent evaluation coordinators. This provides insight that teachers would be very willing to participate in evaluations where other teachers are in control of the process. This in no way rules out the use of an administrator as evaluation coordinator, but suggests that

there may be some collegial value and professional motivation to choose teachers for the job.

The third rated issue, states that program evaluation provides staff with useful information upon which to make decisions. This statement provides two insights. One is that program evaluation provides useful information, and the other deals with staff making decisions. This concept of collegial decision making has been stressed in the literature examining the utilization of evaluation information (Alkin and Stecher, 1983) and also as an important factor in successful internally controlled evaluations (Love, 1983). Love refers to this type of climate as "participative". Following this, the fourth rated issue provides evidence that staff prefer internal control over external. This reinforces the desire for a collegial, participative process which involves staff directly in making decisions and setting policy.

The last issue rated as important encompasses the need for more information, debate and seminars on program evaluation. This may be interpreted in a variety of ways, including program evaluation as a threatening experience. Given the previous information from subproblem two which indicates a slight threat for formalized evaluations, this may indeed be the reason for the rating. On the other hand, it was suggested that teachers could be competent evaluation coordinators. This statement may have drawn from those who wish more information on processes, techniques and methods. It is likely that both is the case.

Only one subgroup difference was determined to be significant. That involved a gender difference over the statement that recommended changes are rarely implemented. The mean for the male group represented moderate agreement, but the female mean revealed slight disagreement. This difference was significant at the .01 level.

Subproblem Four - Summary

Subproblem four states:

To what extent do the attitudes of staff fall into the categories determined to exist by Turner and Clift (1985) in a similar study of schools in Solihull England?

This section of the study dealt only with the subclassification of program evaluation referred to as internal. Control of internal evaluation remains within the school. No external evaluation coordinator is used. Within this classification are two subclassifications which deal with formality. Those evaluations which involve the use of accepted methods or procedures are classified as formal. Those which follow no set guidelines are classified as informal. In order for realistic comparisons to be made between the findings of the Solihull study and this one, only the results of internal program evaluation were dealt with.

Using formal-internal program evaluation information as the basis of comparison, the staff responses factored into five categories as did those in the Solihull study. Through these five major factors, 62.8% of the variance in attitudes was explained. The major difference was the exclusion of the concept "career advancement" which Turner and Clift (1985) referred to as professional development. Professional development (career advancement) was a separate factor in the Solihull study. In these findings, professional development loaded with professionality, perhaps because of the different connotation which the term possessed. The exclusion of variables which could have pointed to career advancement as a factor in this study was a flaw in the instrument design. The other

minor difference was that the concepts of threat and efficiency created two distinct factors rather than loading together as in the study by Turner and Clift.

Informal program evaluation provided a factorial structure which was unique in several ways when compared with the formal-internal results. Firstly, six factors were determined to account for 52.2% of the variance in attitudes. Second was the creation of an unknown major factor which combined the concepts of efficiency and openness. The occurrence of these two concepts together is not logical unless one envisions a continuum with openness at one end, and threat that the other. This would point to existence of another factor which could account for the loading of efficiency and "lack of threat" together. The remaining five factors covered the same concepts as existed in formal-internal evaluation.

Subproblem Four - Conclusions

The differences which occurred in formal-internal evaluation compared to Solihull were minor in nature. The same basic factors were determined to exist within both studies. The strength of the similarities in the results of the two studies are quite interesting and may begin to point to the general existence of these factors as common to formal-internal evaluations.

The differences which existed between the factors created with informal evaluation and formal-internal lead to speculation about the cause. The difference in the types of evaluation may be a large contributor. Both types are considered to be internal, but but informal follows no set procedures. The methodology which suggests formality may play a large part in not only the factors involved, but also the thoroughness and extent of evaluation which takes place. This infers then that the results of the two types of evaluations could be quite different. As a result, one must then assume that the purpose for which the evaluation was undertaken, may indeed

play a major role in the factorial causation of the attitudes towards each type of evaluation. Based on those assumptions, it can be proposed that the factor which combined efficiency and openness variables may conceptually deal with "practicality".

Informal program evaluation is generally undertaken to determine the modifications needed for a specific unit of study. The changes implemented are primarily based upon personal observation of actual results contrasted with expected results. The process itself is efficient and open. The students are involved either directly or indirectly. Time is not "wasted" in gathering data from a variety of sources. Changes are usually implemented quickly. Practicality is the centre of the process. The need for change is seen, and changes are implemented to rectify the perceived problem.

To make ongoing use of program evaluation within the current school structure, it must be practical. This would expect that thoroughness would be sacrificed in an effort to keep within reasonable time lines. Stated another way, informal program evaluation may be used to assess a part of a program in an isolated way. It may be used to assess the extent to which short term program goals are reached, or even be intertwined with needs assessment.

Informal program evaluation is normally undertaken by a single teacher for the purposes of reaching decisions that will have impact only within the isolation of the classroom. A process can be very efficient of time if while planning, a teacher revises a learning unit to overcome previous difficulties. Openness would be natural, as the changes are for the benefit of the students, and students may even be involved in the suggestion of changes (likes and dislikes). All of the above can be

described in terms of the concept practicality. This would not be a concept expected to be associated with formal types of program evaluation.

Implications and Recommendations

This study was prompted to some extent by the circulation of the document Indicators of Effective Programs (1989) within the Calgary Board of Education. The outcome of this study has implications for the system-wide implementation of this program or any other plan for the evaluation of programs. As well there are implications for the field of program evaluation. The following is a discussion of the implications and recommendations which would assist in further understanding of program evaluation usage in schools.

Educational Implications

It is clear that teachers feel the process of program evaluation is valuable and an activity which should be undertaken as professionals. It is generally thought of as a process which stimulates new ideas, pinpoints strengths as well as weaknesses, can be a source of motivation and can promote change in schools.

The strongest of the implementation issues dealt with teachers feeling the need to benefit in effectiveness as a result of participation in program evaluation. This reflects on the statement by Smith (1981a) that the result must provide "relevant, useful, timely information"(p. 79) for teachers to become more effective. This alone implies their collective belief that when program evaluation is used, it should be for the purposes of improvement.

In addition participants preference for evaluation control was internal. Teachers feel that internal evaluation is not likely to be superficial, and that this form of evaluation is not a waste of time. Additionally, the feeling of threat is limited for evaluations with internal control. With formal-internal evaluation, teachers deemed that the use of questionnaires and written reports were of value. This indicates a commitment to obtaining others opinions, and sharing the results. Love (1983, p. 8) suggests that "internal evaluation systems are intended to influence decision making and organizational performance". One of the drawbacks of internal evaluation, is that teachers rarely see the total school as an organization having an impact on students. Most frequently, teachers reveal their domain as bounded by the walls of their classroom. There is a sense of isolation in the work they do with students. Therefore it is important that if formal-internal evaluation is undertaken, it is as a total school activity.

Informal evaluation shares all of the positive characteristics displayed by formal-internal evaluation, but without the questionnaires and written report. This tends to indicate that the process is used for isolated purposes, that the results are not readily shared and that opinions other than those of students are not sought. The use of informal program evaluation is not likely to have an impact on the total organization, but is still felt to be of value.

The use of checklists were rated as slightly agree, indicative of only having slight value. This is most likely because checklists rarely involve the school-based issues which are a result of the culture and organization of the school. This has important implications for the Indicators of Effective Programs as proposed by the Program Services Centre of the Calgary Board of Education. Although the intent of this document was to "guide" in the development of school-based indicators of

effective "curriculum programs", its authors included complete checklists developed by curriculum specialists. In addition, the suggested use of these indicators was to guide the improvement of programs within the school, without providing a method. The implied method was to measure the extent to which school programs met the standards of the curriculum experts.

Administrators should realize that teachers are not in favor of using student scores as a measure of program effectiveness. Teachers are also not impressed with the thought of administrators being sole evaluators of programs. These two methodologies are unlikely to be widely used by rational administrators, but together they illustrate two conditions which would be considered threatening to teachers.

Although by definition in this study, program evaluation does not include teacher evaluation, it is of great interest that observing other teachers' lessons, and coupling the process with self-evaluation were rated as items to be included. This may be because "teaching" and "program" are carefully intertwined. The interconnection of the activity with the process is what contributes in part to the effectiveness. Teachers are willing to examine this connection, but only in non-threatening ways. This is substantiated by the mean values for observing other's lessons. For evaluations with internal control, the values are 3.8 and 3.7, but where the control becomes external, the willingness to pursue this activity drops to 2.9. Even with external control, teachers are still willing to self-evaluate, assuming that they would disclose just what they were comfortable with.

With formal evaluations, teachers feel that collecting parents views are important, but more important where the control of the evaluation is external. This may be a reflection on a perceived difference between evaluations with internal

control verses those with external control. It may be implied that parents views are felt to be more valuable for purposes of accountability rather than improvement.

The factors which accounted for 62.8% of the variance in attitudes for formal-internal evaluation were made up of professionality / professional development, efficiency, threat, collegiality / openness and formality. These are the major conceptual issues which were found to contribute to how teachers feel about program evaluation. Careful consideration must be given to each of these factors when developing or implementing a plan of program evaluation. Informal program evaluation also includes these factor plus one other, termed in this study as practicality for lack of a better description.

Research Implications

Implications of this study for the discipline of program evaluation are primarily those which support many constructs and ideas which have been published in research literature. The subdivision of program evaluation classification by formality and locus of control enables the drawing of distinctions among them. Most researchers in the field of program evaluation neglect or ignore the existence of informal evaluation and concentrate on formal evaluation. By comparing attitudes of teachers based on experience in both formal-internal and formal-external evaluations, it is clear that teachers are more supportive of the ability of formal-internal evaluation to provide useful information and promote school change than formal-external evaluation.

Finding the existence of five factors which represented the same conceptual base as those of the Solihull study by Turner and Clift (1985) provide reason to surmise that these factor may be involved in this type of program evaluation, in

other populations. It indicates that further research is needed with a more varied population to confirm this trend.

The determination of certain issues to be of prime importance to teachers involved in program evaluation lead to the question of how wide spread these attitudes are. Further studies of broader scope are needed to determine this. In addition, given the benefits which staff perceive to follow involvement in program evaluation, further study is needed to determine if such change actually takes place as a result of formal-internal program evaluation. A study titled Elementary School Self-Evaluation and Curriculum Change by Roger Webb (1971) suggest that in fact change does occur during the year of formal-internal evaluation. No follow-up was done to determine if change continued in successive years.

Recommendations

Recommendations which arise as a result of this study are most applicable to the population from which the sample was taken. Generalization of these results to other populations must be made only with careful consideration and reasoned thought, as there is little data to link these attitudes to a broader base of teachers. The similarity between factors identified within these attitudes and those in the Solihull study provide an indication that the findings may be applicable to other populations, but is in no way definitive. The recommendations derived from the views of this sample of teachers are divided into two categories. First the recommendations which apply specifically to the Calgary Board of Education and may have some general usefulness for education systems in Alberta, followed by those for the discipline of program evaluation.

Education. The first recommendation for education is that the continuation of informal program evaluation be encouraged, but that the organizational

administration realize the impact will be limited to fine tuning and revision at the classroom level. The thoroughness of this process is teacher dependent, but is most often limited in scope.

Second, the use of resources to support evaluations with external control should be limited. Teachers express the opinion that this form of evaluation is the most likely of the three types, to be superficial in nature. Its ability to act as a positive growth tool is also much less than the other types. Therefore it is recommended that evaluations with internal control be used where the intent is improvement of education.

Third, the value of formal-internal program evaluation should be recognized, and the positive attitudes which exist in relation to its ability to motivate and promote change within the school.

Fourth, that wherever possible, teachers be placed in the role of evaluation coordinators. Prior to this taking place, inservice training should be provided to interested teachers related specifically to methods which are useful for internal program evaluation. Methods should be tailored to being responsive to school needs, involving stakeholders in every stage of the process and understanding the limited usefulness of checklists. Flexible staffing arrangements should be in place to accommodate for the necessary time commitment and work load.

Research. Further research is needed into all aspects of informal program evaluation as a method for program improvement. Specifically, research must be conducted into the potential underlying factors associated with teacher attitudes as they pertain to informal evaluation. Further research is needed to determine if the existence of six attitude factors occurs again for informal evaluation, and indeed if

the new factor termed in this study as "practicality" is a logical outgrowth of these attitudes.

Second, further research is needed to determine if "career enhancement" is a major factor in the attitudes of staff to formal-internal program evaluation.

Third, further research is needed to create a useful guidebook which will assist evaluation coordinators to conduct internal evaluations which meet their needs for information. This guidebook must minimize the use of pre-prepared checklists, and encourage a total school effort in the study.

Fourth, further research is need to determine the extent to which the five basic factors which determine attitudes toward formal-internal evaluation, exist in a broader population.

Bibliography

Adelman, C. & Alexander, R. F. (1982). The self-evaluating institution: Practice and principles in the management of educational change. New York: Methuen & Co.

Alberta Teachers' Association. (1969): Self-evaluation: A guide for the elementary school. Edmonton: Author.

Alberta Teachers' Association. (1989). School and program evaluation: A manual for teachers. Edmonton: Author.

Alkin, M. C., Daillak, R. & White, P. (1979). Using evaluations: Does evaluation make a difference? Beverly Hills: SAGE Publications, Inc.

Alkin, M. C. & Stecher. B. (1983). Evaluation in context: Information use in elementary school decision making. Studies in Educational Evaluation, 9(1), 23-32.

Attkisson, C. C. & Broskowski, A. (1978). Evaluation and the emerging human service concept. In C. C. Attkisson, W.A. Hargreaves, M. J. Horowitz, & J. E. Sorensen (Eds.). Evaluation of human service programs. New York: Academic Press.

Ayers, J. B. (1986). Design characteristics for meaningful program evaluation. Cookeville, TN: Tennessee Technological University.

Brewer, M. B. (1983). Evaluation: Past and present. In E. L. Struening & M. B. Brewer (Eds.). Handbook of evaluation research. University edition. Beverly Hills: SAGE Publications, Inc.

Burgess, J. O. (Ed.). (1987). School-based program evaluation manual. Regina, Saskatchewan: Saskatchewan Instructional Development and Research Unit.

Calgary Board of Education. (1989). Indicators of effective programs. Unpublished document developed by Educational Leadership Centre Program Services. Calgary: Author.

California State Department. (1987). Program quality review for elementary schools: Process, criteria, and self-study. Sacramento: Author.

California State Department. (1987). Program quality review for high schools: Process, criteria, and self-study. Sacramento: Author.

Campbell, D. T. (1979). Assessing the impact of planned social change. Evaluation and Program Planning, 2, 67-90.

Clift, P. S., Nuttall, D. L. & McCormick, R. (Eds.). (1987). Studies in school self-evaluation. London: The Falmer Press.

DeRoche, E. F. (1987). An administrator's guide for evaluating programs and personnel: An effective schools approach. Second Edition. Boston: Allyn and Bacon, Inc.

Ewell, P. (1984). The self-regarding institution: Information for excellence. NCHEMS Executive Overview. Boulder, Colerado: National Center for Higher Education Management Systems.

Greene, J. G. (1988) Stakeholder participation and utilization in program evaluation. Evaluation Review, 12(2), 91-116.

Guba, E. G. (1978). Toward a methodology of naturalistic inquiry in educational evaluation. C. S. E. Monograph Series in Evaluation, No. 8. Los Angeles: Center for the Study of Evaluation, University of California.

Herman, J. L. (1987). Evaluation for school improvement: Try-out of a comprehensive school based model. Los Angeles: Center for the Study of Evaluation.

Johnson, C.F. & Christal, M.E. (1985). Preparing for self-study. The AIR Professional File, 22.

Love, A. J. (1983). The organizational context and the development of internal evaluation. In A. J. Love (Ed.). Developing Effective Internal Evaluation. New Directions for Program evaluation, No. 20. San Francisco: Jossey-Bass.

Meyers, W. R. (1981). The evaluation enterprise: A realistic appraisal of evaluation careers, methods, and applications. San Francisco: Jossey-Bass Inc., Publishers.

National Commission on Excellence in Education. (1983). A nation at risk: The imperative for educational reform: A report to the nation and the secretary of education. United States Department of Education.

Newman, D. L, Brown, R. D., & Rivers, L. S. (1983). Locus of control and evaluation use: Does sense of control affect information needs and decision making? Studies in Educational Evaluation, 9, 77-88.

Parlett, M., & Hamilton, D. (1976). Evaluation as illumination: A new approach to the study of innovative programs. In G. V. Glass (Ed.). Evaluation studies review annual, 1. Beverly Hills: SAGE Publications, Inc.

Rein, M. (1981). Comprehensive program evaluation. In R. Lavine, M. Solomon, G. Hellstern & H. Wollmann (Eds.). Evaluation research and practice: Comparative and international perspectives. Beverly Hills: SAGE Publications, Inc.

Rossi, P. H., & Wright, S. R. (1977). Evaluation research: An evaluation of theory,practice, and politics, Evaluation Quarterly, 1.

Smith, N. L. (1981). Creating alternative methods for educational evaluation. In N.L. Smith (Ed.), Federal efforts to develop new evaluation methods (pp. 77-94). San Francisco: Jossey-Bass Inc., Publishers.

Smith, N.L. (1981). Developing evaluation methods. In N. L. Smith (ed.), Metaphors for evaluation: Sources of new methods (pp. 17-49). Beverly Hills: SAGE Publications, Inc.

Stake, R. E. (1975). Program evaluation: Particularly responsive evaluation. Occasional Paper No. 5. Kalamazoo: Evaluation Center, Western Michigan University.

Stake, R. E. (1983). Stakeholder influence in the evaluation of cities-in schools. New Directions in Program Evaluation, 17, 15-30.

Stufflebeam, D. L. (1971). Educational evaluation and decision making. Ithica: Peacock Inc.

Stufflebeam, D. L. & Shinkfield, A. J. (1985). Systematic evaluation. Boston: Kluwer-Nijhoff Publishing.

Vogt, J. F. & Hunt, B.D. (1988). What really goes wrong with participative work groups? Training and Development Journal, 42(5), 96-100.

Webb, R. (1971). Elementary school self-evaluation and curriculum change. New Jersey: Educational Technology Publications.

West Vancouver SD. #45. (1986). Self-evaluation for elementary schools. Vancouver: Author.

Windele, C. & Neigher, W. (1978). Ethical problems in program evaluation: Advice for trapped evaluators. Evaluation and Program Planning, 1(2), 97-107.

Appendix 1

Instrument Cover Letter

«DATA sample data base.word»
May 20, 1990

«full name»
«SCHOOL»

Dear «first name»:

I am currently undertaking graduate research as part of the requirements for a masters degree at the University of Alberta. My research involves determining staff attitudes towards "program evaluation", in order to determine if "program evaluation" can become a useful tool in the school improvement process.

You have been randomly selected to participate in this study but are under no obligation to complete this questionnaire. Your assistance in filling out and returning the questionnaire will be of great value to me in completing this study. Your responses will remain anonymous. The number on the return envelope will be used only to follow up non-respondents. Records of the name-number match will be strictly guarded to assure your protection.

I have attempted to create an instrument which will take less than 15 minutes to complete. I would appreciate the donation of your time, to assist me in gaining a greater insight into the practice of program evaluation within the Calgary Board of Education. To determine overall attitudes towards this practice your opinions are vital.

If you wish to receive a brief summary of the data collected, please place an "X" in the box provided in the questionnaire.

I thank you in advance, for your cooperation in completing this questionnaire.

Sincerely,

Wayne Morris

Appendix 2

Survey Instrument

Definitions:

Program evaluation: The process of delineating, obtaining and providing useful information for making decisions about a planned sequence of activities, which are intended to achieve some goal.

- does **NOT** include the evaluation of teacher performance.

The **focus** for program evaluation may either be narrow or broad, depending upon the need for information.

Example 1: A school-wide evaluation used to assess a school's overall educational program.

Example 2: A curriculum specific evaluation designed to assess the extent to which implementation is occurring.

Example 3: A non-curriculum evaluation used to assess the effectiveness of a "peer leadership" program in the school.

Program evaluation also goes by the names of **program review** or **program planning**. It may be either formative or summative in nature.

Formal evaluation: is a process that follows or adheres to accepted forms, conventions or regulations. These can involve a group of people, with one member being appointed coordinator of the evaluation. An evaluation design is determined, which includes evaluation procedures, the data to be collected, and the strategies for disseminating information to stakeholders.

Informal evaluation: is frequently a teacher-initiated activity, following no formalized, structured or accepted procedure. It may involve only the teacher, or be expanded to include students and colleagues. Often, no formal report is compiled, but the information is utilized by the teacher, to improve the existing program.

Internal evaluation: takes place within the context of the organization. In this case, the school uses the resources available within its walls, without the services of external subject specialists or consultants. The overall control of the evaluation remains in school hands.

External evaluation: utilizes expertise existing outside the school. Usually, the control of the evaluation is in the hands of these outside experts.

Directions:

Please complete the questionnaire and return it in the self-addressed envelope provided, via **Interschool MAIL.**

❑ Yes! I wish to receive a copy of the results from this study.

A: Please complete the following demographic information:

Place an "X" in the appropriate box.

1. Number of years experience as a teacher and or administrator:
   - ❑ 1 - 5 years
   - ❑ 6 - 10 years
   - ❑ 11 or more years

2. Are you presently in receipt of an administrative allowance?
   - ❑ yes
   - ❑ no

3. Check appropriately:
   - ❑ teacher
   - ❑ principal

4. Check appropriately:
   - ❑ female
   - ❑ male

5. Please check the appropriate level for the school in which you work:
   - ❑ elementary
   - ❑ junior high
   - ❑ senior high

6. Number of times involved in the evaluation of a school-based program
   - ❑ 0
   - ❑ 1 - 2
   - ❑ 3 - 5
   - ❑ 6 - 10
   - ❑ 11 or more

B: Attitudes towards "Informal" Program Evaluation

Please refer to the definitions attached before answering the questions in this section.

Please answer all questions in this section in reference to your experience with **informal** program evaluation.

Use the following key in responding to each of the following statements. Circle the number which best reflects your opinion concerning the statement presented.

Key: SA - Strongly agree, A - agree, U - undecided, D - disagree, SD - Strongly disagree

1. Of all your program evaluations, how many would you classify as being "informal"?

- ❑ 0
- ❑ 1 - 2
- ❑ 3 - 5
- ❑ 6 - 10
- ❑ 11 or more

If you checked "0", proceed to section "C".

Informal program evaluation:

	SA	A	U	D	SD
2. should include teachers observing other's lessons	1	2	3	4	5
3. can promote change in schools	1	2	3	4	5
4. should involve questionnaires	1	2	3	4	5
5. should assess parents' views	1	2	3	4	5
6. tends to make teachers feel threatened	1	2	3	4	5
7. should be coupled with self-evaluation	1	2	3	4	5
8. is, on the whole, a good idea	1	2	3	4	5
9. should be compulsory	1	2	3	4	5
10. should include a written report	1	2	3	4	5
11. is virtually a waste of time	1	2	3	4	5
12. can stimulate new ideas	1	2	3	4	5
13. should utilize student scores from provincial tests	1	2	3	4	5
14. should be done by administrators only	1	2	3	4	5
15. can be a source of motivation for teachers	1	2	3	4	5
16. provides a good basis for questioning existing methods	1	2	3	4	5
17. pinpoints strengths as well as weaknesses	1	2	3	4	5
18. should include taking pupils' views into account	1	2	3	4	5
19. is likely to be superficial	1	2	3	4	5
20. is too time consuming	1	2	3	4	5
21. should involve checklists	1	2	3	4	5

C: Attitudes towards "Formal Internal" Program Evaluation

Please refer to the definitions attached before answering the questions in this section.

Please answer all questions in this section in reference to your experience with **formal internal** program evaluation.

Use the following key in responding to each of the following statements. Circle the number which best reflects your opinion concerning the statement presented.

Key: SA - Strongly agree, A - agree, U - undecided, D - disagree, SD - Strongly disagree

1. Of all your program evaluations, how many would you classify as being "formal internal"?
   - ❑ 0
   - ❑ 1 - 2
   - ❑ 3 - 5
   - ❑ 6 - 10
   - ❑ 11 or more

If you checked "0", proceed to section "D".

Formal internal program evaluation:

	SA	A	U	D	SD
2. should include teachers observing other's lessons	1	2	3	4	5
3. can promote change in schools	1	2	3	4	5
4. should involve questionnaires	1	2	3	4	5
5. should assess parents' views	1	2	3	4	5
6. tends to make teachers feel threatened	1	2	3	4	5
7. should be coupled with self-evaluation	1	2	3	4	5
8. is, on the whole, a good idea	1	2	3	4	5
9. should be compulsory	1	2	3	4	5
10. should include a written report	1	2	3	4	5
11. is virtually a waste of time	1	2	3	4	5
12. can stimulate new ideas	1	2	3	4	5
13. should utilize student scores from provincial tests	1	2	3	4	5
14. should be done by administrators only	1	2	3	4	5
15. can be a source of motivation for teachers	1	2	3	4	5
16. provides a good basis for questioning existing methods	1	2	3	4	5
17. pinpoints strengths as well as weaknesses	1	2	3	4	5
18. should include taking pupils' views into account	1	2	3	4	5
19. is likely to be superficial	1	2	3	4	5
20. is too time consuming	1	2	3	4	5
21. should involve checklists	1	2	3	4	5

D: Attitudes towards "Formal External" Program Evaluation

Please refer to the definitions attached before answering the questions in this section.

Please answer all questions in this section in reference to your experience with **formal external** program evaluation.

Use the following key in responding to each of the following statements. Circle the number which best reflects your opinion concerning the statement presented.

Key: SA - Strongly agree, A - agree, U - undecided, D - disagree, SD - Strongly disagree

1. Of all your program evaluations, how many would you classify as being "formal external"?

- ❑ 0
- ❑ 1 - 2
- ❑ 3 - 5
- ❑ 6 - 10
- ❑ 11 or more

If you checked "0", proceed to section "E".

Formal-external program evaluation:

	SA	A	U	D	SD
2. should include teachers observing other's lessons	1	2	3	4	5
3. can promote change in schools	1	2	3	4	5
4. should involve questionnaires	1	2	3	4	5
5. should assess parents' views	1	2	3	4	5
6. tends to make teachers feel threatened	1	2	3	4	5
7. should be coupled with self-evaluation	1	2	3	4	5
8. is, on the whole, a good idea	1	2	3	4	5
9. should be compulsory	1	2	3	4	5
10. should include a written report	1	2	3	4	5
11. is virtually a waste of time	1	2	3	4	5
12. can stimulate new ideas	1	2	3	4	5
13. should utilize student scores from provincial tests	1	2	3	4	5
14. should be done by administrators only	1	2	3	4	5
15. can be a source of motivation for teachers	1	2	3	4	5
16. provides a good basis for questioning existing methods	1	2	3	4	5
17. pinpoints strengths as well as weaknesses	1	2	3	4	5
18. should include taking pupils' views into account	1	2	3	4	5
19. is likely to be superficial	1	2	3	4	5
20. is too time consuming	1	2	3	4	5
21. should involve checklists	1	2	3	4	5

E: General Impressions of Program Evaluation

Please refer to the definitions attached before answering the questions in this section.

Use the following key in responding to each of the following statements. Circle the number which best reflects your opinion concerning the statement presented.

Key: SA - Strongly agree, A - agree, U - undecided, D - disagree, SD - Strongly disagree

	SA	A	U	D	SD
1. Recommended changes are rarely implemented	1	2	3	4	5
2. Teachers can be competent evaluation coordinators	1	2	3	4	5
3. Program evaluation provides staff with useful information upon which to make decisions	1	2	3	4	5
4. Alberta Education should mandate program evaluation	1	2	3	4	5
5. Teachers need to feel that their effectiveness will be improved in some way as a result of being involved in program evaluation	1	2	3	4	5
6. More information, debate and seminars on program evaluation are required	1	2	3	4	5
7. School-based control of program evaluation is preferable to outside control	1	2	3	4	5

Appendix 3

Rotated Factor Loadings

Rotated Factor Loadings
Informal Program Evaluation

Factors	1	2	3	4	5	6	7
Variable							
B(10)	-0.81	-0.02	-0.11	-0.09	-0.04	-0.16	0.15
B(11)	0.71	0.16	0.09	0.10	0.18	0.13	0.08
B(19)	-0.60	0.10	-0.00	-0.42	-0.10	0.11	-0.07
B(17)	0.55	0.19	0.38	0.20	0.00	0.18	-0.22
B(5)	-0.55	0.14	0.23	-0.20	-0.12	0.16	-0.26
B(13)	-0.53	-0.34	0.03	-0.08	-0.43	0.12	0.05
B(12)	-0.08	-0.80	0.12	-0.11	0.08	-0.09	-0.13
E(7)	-0.04	-0.73	-0.04	0.01	-0.03	0.25	0.09
B(4)	0.25	-0.05	0.81	-0.07	0.03	-0.10	-0.02
B(9)	-0.10	-0.09	0.75	0.18	0.04	-0.23	-0.08
B(15)	0.21	0.08	0.04	0.81	0.07	-0.03	-0.18
B(14)	0.12	0.22	0.19	0.63	0.26	0.27	0.04
B(18)	-0.37	0.10	0.07	-0.63	0.10	-0.23	0.05
B(16)	0.13	0.05	0.06	0.57	0.28	0.05	-0.34
B(8)	0.07	-0.05	0.29	0.51	0.05	0.48	0.30
E(5)	-0.17	0.03	-0.02	-0.00	-0.84	-0.15	0.11
B(2)	0.24	-0.20	-0.01	0.32	0.64	0.26	-0.10
E(4)	-0.08	0.18	-0.11	-0.12	0.00	-0.78	0.02
B(1)	0.05	0.03	-0.05	0.07	0.26	0.78	-0.17
E(6)	0.01	0.05	-0.08	-0.22	-0.11	-0.09	0.84
B(6)	0.02	0.14	0.39	0.30	0.47	-0.10	0.27
B(7)	0.47	0.05	0.20	0.33	0.33	0.30	-0.00

Variance Explained by Rotated Components

1	2	3	4	5	6	7
3.01	1.56	1.80	2.73	1.93	2.04	1.26

Percent of Total Variance Explained

1	2	3	4	5	6	7
13.67	7.08	8.17	12.39	8.79	9.27	5.73

Rotated Factor Loadings
Formal-Internal Program Evaluation

Factors	1	2	3	4	5
Variable					
C(14)	0.85	0.26	-0.06	0.10	0.10
C(15)	0.84	0.30	-0.01	0.17	0.03
C(11)	0.80	0.07	-0.05	0.17	0.05
C(2)	0.64	0.20	0.08	0.50	-0.10
C(7)	0.58	0.22	0.00	0.06	0.39
C(16)	0.53	0.27	0.16	0.44	0.39
C(18)	-0.48	-0.74	-0.03	0.10	-0.08
C(19)	-0.19	-0.74	0.10	0.03	-0.09
C(10)	-0.35	-0.71	-0.19	-0.08	-0.26
E(6)	-0.15	-0.68	0.11	-0.20	-0.19
C(5)	-0.38	0.13	0.73	-0.03	-0.10
C(12)	-0.11	0.28	-0.69	0.01	0.03
C(6)	0.50	0.13	0.54	-0.05	0.23
C(4)	0.26	-0.09	-0.07	0.66	0.11
C(17)	0.42	0.04	-0.01	0.61	0.22
C(1)	0.04	0.47	0.22	0.57	-0.18
C(13)	-0.34	-0.18	0.22	-0.55	0.41
C(8)	0.10	0.08	-0.00	0.08	0.82
E(4)	-0.07	-0.44	0.06	0.07	-0.59
C(9)	0.22	0.24	0.48	0.20	0.50
E(5)	0.03	-0.29	0.11	-0.32	-0.26
E(7)	0.23	0.15	-0.31	-0.40	-0.04

Variance Explained by Rotated Components

1	2	3	4	5
4.39	3.13	1.84	2.34	2.12

Percent of Total Variance Explained

1	2	3	4	5
19.94	14.24	8.36	10.65	9.62